SciDAC Visualization and Analytics Center for Enabling Technologies
Semi-Annual Progress Report
May 2010 through October 2010

E. Wes Bethel[*] and Chris Johnson[†]
Principal Investigators

Charles Hansen, Valerio Pascucci, Claudio Silva, Allen Sanderson, Tom Fogal, Kristi Potter, Attila Gyulassy, Brian Summa, Josh Lavine, Sidarth Kumar, Giorgo Scorzelli, Abhishek Tripathi, Julien Tierney, Lauro Lins, Juliana Freire, Emanuele Santos, Harsh Bhatia, Guoning Chen, Shreeraj Jadhav, Luis Nonato [‡]

Sean Ahern, George Ostrouchov, Dave Pugmire, Jeremy Meredith[§]

Eric Brugger, Peer-Timo Bremer, Brad Whitlock[¶]

Ken Joy, Eduard Deines, Christoph Garth, David Camp, Bernd Hamann[‖]

David Camp, Hank Childs, Prabhat, Oliver Rübel, Daniela Ushizima, Gunther Weber[**]

November 2010

[*] Lawrence Berkeley National Laboratory
[†] Scientific Computing Institute, University of Utah
[‡] Scientific Computing Institute, University of Utah
[§] Oak Ridge National Laboratory
[¶] Lawrence Livermore National Laboratory
[‖] University of California – Davis
[**] Lawrence Berkeley National Laboratory

Contents

1 Executive Summary

1.1 Overview

The SciDAC Visualization and Analytics Center for Enabling Technologies (VACET) focuses on leveraging scientific visualization and analytics software technology as an enabling technology for increasing scientific productivity and insight. Our mission is to foster scientific insight through creating and deploying effective data understanding technology that is truly responsive to the needs of our stakeholders in the scientific research community who are "awash in data." It is widely accepted that one of the bottlenecks in contemporary science is the need to gain insight from vast collections of complex data.

The vision for our Center is to respond directly to this challenge by adapting, extending, creating when necessary and deploying visualization and data understanding technologies for our science stakeholders. Organized as a Center for Enabling Technologies, we are well positioned to be responsive to the needs of a diverse set of scientific stakeholders in a coordinated fashion using a range of visualization, mathematics, statistics, computer and computational science and data management technologies.

We are pleased to report accomplishments during the period of December 2009 through May 2010, both in terms of impact for scientific stakeholders and in terms of providing leadership in the visualization and analysis community.

1.2 Accomplishments

Science Application Projects

Astrophysics. One of VACET's central mission objectives is to provide production-quality, petascale-capable visual data analysis software to the science community. To that end, many of the researchers in the SciDAC Community Astrophysics Consoritium (CAC) have adopted VisIt for project-wide use. This accomplishment is facilitated by ongoing VACET efforts to enhance VisIt to meet CAC needs and to help them apply VisIt to challenging science problems. Recent accomplishments include production of a movie that won a People's Choice Award at the SciDAC 2010 Visualization Night, images that are featured on Time's Health & Science website, as well as numerous science-related press releases (Section 2.1). Related, we are achieiving similar objectives with another astrophysics research group, T. Mezzacappa. There, our team is providing the software infrastructure and expertise in its use to aid in studying how magnetic fields play a role in supernovae explosions (Section 2.2).

Accelerator. As part of our long-standing relationship with the SciDAC Community Petascale Project for Accelerator Science and Simulation (COMPASS), we have begun activities aimed at helping the team at Fermi that uses the Synergia code make the transition to VisIt. As this code is run at higher spatiotemporal resolution, existing legacy analysis solutions are incapable of processing the ever-larger data. VACET is helping this team transition to VisIt to meet their visual data analysis needs (Section 2.3).

Groundwater and Environmental Management. In this reporting period, our team has engaged with two different teams. Working with the PFLOTRAN code team, we generated a scalable PFLOTRAN data loader and included it as part of the VisIt distribution (Section 2.4). We have also engaged with the DOE-funded project Advanced Simulation Capability for Environmental Management (ASCEM) to assist them in project-wide adoption of VisIt (Section 2.5).

Fusion. In a collaborative effort with fusion scientists, our team has developed new tools that enable interactive analysis of magnetic islands. This new capability is delivered to the science

community in petascale-capable, production-quality software (Section 2.7). We are also completing implementation of software infrastructure (Section 4.1) that we are deploying to fusion researchers interested in doing more in-depth visual data analysis and exploration of large, particle-based datasets (Section 2.6).

Nuclear Science. We have provided one-on-one consulting assistance to the University Nuclear Energy Density Functional project (UNEDF) to help them use VisIt to study nuclear science problems (Section 2.8).

Technology Incubation Projects

Flow visualization. We are able to capitalize upon our abilities as a Center to create and deploy technology that has far-reaching impact. One of our efforts focuses upon developing robust and highly scalable algorithms for computing integral curves. This code forms the basis for many different visualization and analysis algorithms: streamlines, stream surfaces, pathlines, path surfaces, and Poincaré analysis. This work, which involves many VACET members, is deployed in petascale-capable, production-quality form and is distributed to the worldwide scientific community (Section 3.1).

Hybrid parallelism and volume rendering. Future computational platforms will be constructed using processors having tens to hundreds of cores. Existing MPI-based codes will likely not scale well on such platforms. Our team has investigated using "hybrid-parallelism," which combines traditional MPI-based distributed memory parallel programming with shared-memory programming constructs to evaluate scalability of volume rendering, a staple visualization algorithm. The results, which include scaling runs at the highest level of concurrency ever published in the visualization community, indicate that the hybrid parallel approach runs faster, uses less memory, and consumes less communication bandwidth than traditional MPI-based approaches. This work helps pave the way for future visual data analysis applications to effectively leverage exascale class platforms (Section 3.2).

Uncertainty visualization. Uncertainty, in the form of confidence, variability, and error, as well as model bias and trends, is used to express descriptive, qualitative characteristics of the data. Because uncertainty is crucial in understanding the reliability of information and thus in objectives such as decision making, its absence can lead to misrepresentations and incorrect conclusions. However, visual portrayal of uncertainty is a challenging task. Our team is exploring a technique known as the Summary Plot as a means to provide an effective visual portrayal of uncertainty information (Section 3.3).

Edge Maps. As an alternative to traditional numerical integration as the basis for vector field visual data analysis and exploration, our team is exploring an alternative approach called "edge maps." This approach uses piecewise linear interpolation along edges connecting vertices in a computational domain, and provides the means to compute and store numerical error, which aims to produce more informative visual representations of data (Section 3.4).

Parallel Morse-Smale Computation. As part of a strategy to perform analysis concurrent with the simulation to avoid I/O costs, our team is exploring how to best parallelize the computation of the Morse-Smale Complex, which is the basis for a class of topological analysis algorithms (Section 3.5).

Remote Collaborative Visualization Infrastructure. A lightweight viewer-side API facilitates state synchronization amongst multiple, distributed visualization/viewer clients to enable remote, collaborative interaction (Section 3.6).

Software Engineering and Infrastructure

VisIt: Production-quality Data Subsetting. An ongoing effort within VACET is to provide the ability for users to quickly sift through very large data to isolate, visualize and analyze "interesting" data. This work continues to advance and is now deployed in VisIt in production-quality form. A user can select a subset of data via a multivariate range query specified through a parallel coordinates plot, then have this subset automatically visualized and analyzed by active plots and operators within VisIt. The new work supports time-varying particle-based datasets, where particles may come and go over time. One user independently made a striking image and offers his enthusiastic support for our work (Section 4.1).

VisIt: Software Infrastructure for Multi-resolution. VisIt has had the notion of support for Adaptive Mesh Refinement data for some time. We have generalized this capability by providing the ability for multi-resolution operation in cases where a dataset loader can present multiple resolutions of data. The idea is to provide a path forward for visual data analysis and exploration for reduced resolution versions of large data for the purposes of interactivity, for use on resource-constrained platforms, and for planning more resource-intensive analysis on full resolution data. A collaborator at the Swiss National Supercomputer Center offers his endorsement for this approach, along with example images showing results (Section 4.2).

VisIt: Software releases and release engineering. The VisIt project has taken steps to improve its efficiency, openness, and to increase the frequency of releases. The team has migrated to a new bug and feature tracking system to better monitor bugs and feature requests, has made this system available to the worldwide community, and has moved to more frequent production releases to more quickly put new capabilities into the hands of scientists (Section 4.3).

Data-parallel Statistical Analysis. An ongoing effort in VACET is to bring to bear the power of the R statistical analysis package, which is effectively limited to operation on a single core, on large-data problems and on DOE's large parallel computational platforms. To achieve that objective our team is performing software engineering to the Rmpi package to enable use of R in a data-parallel mode (Section 4.4).

ViSUS: Parallel I/O, Redesigned Viewers. The ViSUS library, which forms the core of some of VACET's topological analysis and data exploration software, has been to date primarily a serial software tool. Recent work explores methods for doing parallel I/O directly from codes, and early results show promise in terms of achieving a large fraction of peak on the BG/P platform at ANL. The ViSUS viewer has been redesigned to better accommodate new features in the future (Section 4.5).

2 Specific Stakeholder Projects

2.1 Astrophysics – Computatonal Astrophysics Consortium

The Problem. To ensure that the visualization and analysis needs of the astrophysics community, and specifically the Computational Astrophysics Consortium (CAC), are met. This involves (i) providing them tools for "bread-and-butter" functionality (i.e., capabilities that are "routine"), (ii) providing support, (iii) helping with high end movies, and (iv) helping with high-end analysis.

The Solution. VACET is providing and deploying to the CAC production quality visualization tools to meet day-to-day needs. We are also providing support to the CAC by fielding questions that come up in solving specific science problems, as well as providing in-depth consulting for creating "difficult" images and movies as well as helping to devise new methods for high-end analysis.

The Impact. We are successfully delivering a production-quality visual data exploration and

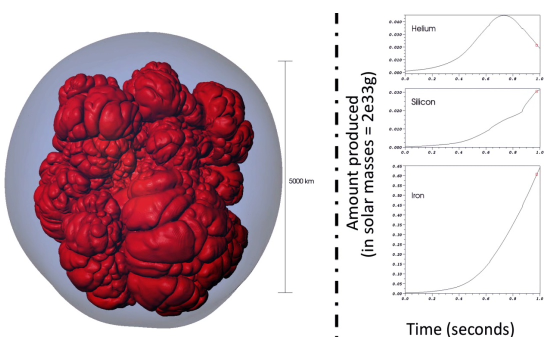
analysis tool to the SciDAC astrophysics community, thereby enabling new ways of exploring and understanding scientific data as well as enabling cost savings through reduced duplication of effort (they don't have to build or buy a tool to meet their visualization needs). Visualization results have been used by CAC members in presentations and publications that describe new science. The visual results are an integral part of both scientific discovery as well as communicating results to the public, other scientists, and stakeholders.

Accomplishments

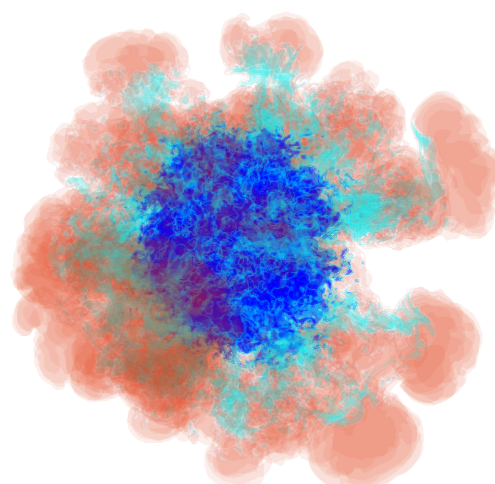
Analysis. We have spent quite a bit of time doing analysis that is powered by our existing particle advection code. Our collaborators (Burrows and Nordhaus, Princeton University) have asked that we keep the details of this analysis private until the results are released. This activity has been championed by David Camp.

Movies and Images.

There have been two movie making efforts. One movie, by Hank Childs, was a winner of a People's Choice award at SciDAC 2010. That movie shows a time-evolving Type Ia supernova explosion computed by CAC researchers using the CASTRO code. The movie can be found on the web[1], and a single frame of the movie is shown below in Figure 1(a). An additional visual example of CASTRO output appears in Figure 1(b). A press release describing scientific results appears on the web[2].



(a) This image comes from a SciDAC 2010 Visualization Night People's Choice award winning movie. It shows a flame front (red), surrounded by the star boundary (blue).



(b) A volume rendering of CASTRO simulation output showing the elements created during a Type Ia supernova. Helium is blue, silicon is cyan, iron is red.

Figure 1: Example visualizations of TypeIa supernova explosion simulations computed by the SciDAC Computational Astrophysics Consortium using the CASTRO code.

A second movie, which is currently under development by David Camp (LBNL), visualizes a simulated Type II supernova explosion computed by CAC researchers using the MAESTRO code. Several frames from this movie are shown in Figure 2. One of these images, the rightmost in Figure 2, was featured on Time's Health & Science website[3].

[1] `http://vis.lbl.gov/~hrchilds/vn_cac.mp4`

[2] http://www.lbl.gov/cs/Archive/news091610.html

[3] `http://www.time.com/time/health/article/0,8599,2021122,00.html`

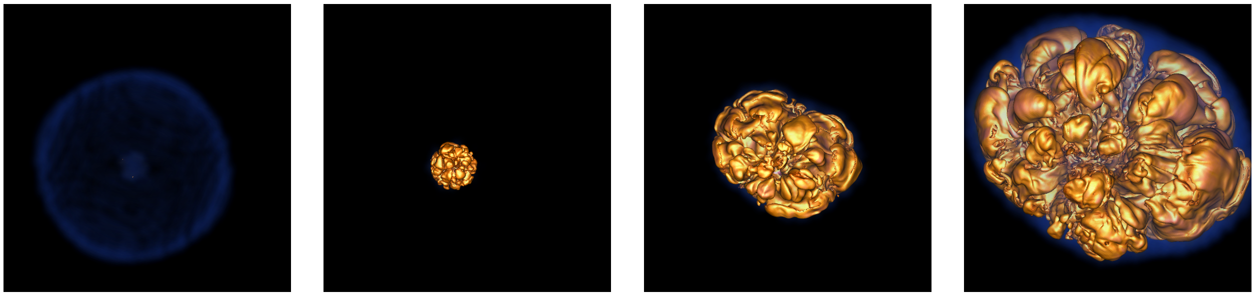


Figure 2: Example visualizations of Type II supernova explosion simulations computed by CAC researchers using the MAESTRO code. The right most image was featured on the Time Health & Science website. These images all show the computed entropy field.

Community-wide Visualization Software

Realizing the vision and mission objective of delivering a production-quality visualization tool to a community of scientific researchers requires ongoing attention to fixing bugs and adding enhancements that enable and accelerate scientific discovery. The that end, We have provided a lot of outreach and support to CAC:

- We have implemented pathlines in VisIt.
- Gunther Weber gave a talk at the CAC all-hands meeting.
- We have participated in a half dozen meetings where we have provided expertise on how to use VisIt to CAC researchers.
- We also have helped UCSC participants set up and use VisIt at both UCSC, SLAC, and Pleides (the NASA machine).
- The CAC asked for the SciDAC 2010 Visualization Night movie to be something they can make in an automated fashion. This movie required a lot of manual effort to combine multiple images, animate curves, etc. In response to this request, Hank added advanced moviemaking features that allow for these features (animate curves, combining multiple windows in an arbitrary way) to be done fully in VisIt. Restated, CAC wouldn't have been able to make the SciDAC 2010 Visaulization Night movie themselves before these enhancements, but they can now after enhancements added to VisIt.
- Other miscellaneous bug fixes in support of CAC:
  - Implemented expression for Shizuka Akiyama to look at position.
  - Implemented Shizuka's suggestion for storing and loading expressions.
  - Fixed problem with VisIt's Cosmos++ reader with reporting time. (Cosmos++ is a code by Jay Salmonson out of LLNL).
  - Provided Shizuka with a script to calculate average density as a function of radius.
  - Numerous other, even-smaller-in-scope support tasks.

Many members of the CAC use VisIt on a regular basis, including researchers from John Bell's group, Stan Woosley's group, and Adam Burrows' group. We have realized the objective of providing a production-quality, petascale-capable visualization tool to this community that meets "bread-and-butter," or routine needs, and that can run on the world's largest computational platforms to perform visual data analysis and exploration of very large data.

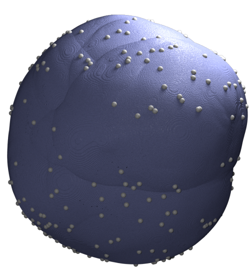
Visualization is crucial. Otherwise, all you have is merely a jumble of numbers. It allows one to diagnose the dynamics, so that the event is not only visualized, but understood. –Adam Burrows, Princeton University

2.2 Astrophysics – T. Mezzacappa

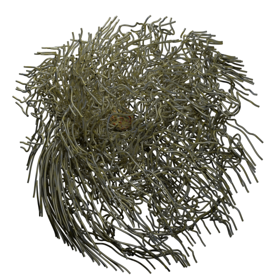
The Problem. Astrophysics code teams have a diverse set of analysis and visualization needs. We are working with these teams to provide them production quality visualization and analysis software for day-to-day use, providing consulting in advanced use of these capabilities, and performing development of new algorithms and techniques that meet new, domain-specific needs. One major need is the ability to explore and analyze magnetic fields in supernovae simulation data. The magnetic fields around the core of a supernova are typically very complicated; gaining deeper understanding of these fields is critical to a better understanding of supernovae explosion mechanics.

The Solution. Our solution is to take a holistic approach that includes diverse activities like software development in the form of algorithm optimizations and feature enhancements, one-on-one assistance in learning to use VisIt, and help creating high-end movies and images for internal and external use. Support has been performed at a variety of levels. For this period, we have implemented a new set of streamline seeding operators that help address the needs of magnetic field analysis, consulted with the astrophysicists in their use, and deployed the new capabilities in VisIt.

The Impact. Particle advection optimizations have resulted in significant speedups for streamline calculations, which in terms accelerates scientific discovery. A new GUI should significantly reduce problems encountered when generating streamline plots. Several code team members have been trained in python scripting for VisIt, and are using these tools to do day-to-day visualization and analysis. Overall, the impact is a combination of faster and better analysis software, increased usability of advanced analysis capabilities, and adoption of VisIt by an astrophysics research group for use in performing critical data analysis activities.



(a) Shock surface shown in blue, seed points shown in white.



(b) Streamlines through magnetic field shown in gold, with the proto-neutron star (PNS) shown in the center.

Figure 3: New features requested by science users result in new capabilities to help study science problems. These new capabilities are distributed to the worldwide science community in VisIt.

Accomplishments.

- A new particle advection GUI is complete, and ready to be deployed.
- New particle seeding options have been delivered in VisIt.
- What was requested was randomized seed points on the surface, and interior of a sphere. These were added to the sphere seeding, as well as all of the other seeding options (Figure 3(a)).
- Our re-factored particle advection code deployed to community in VisIt. This improved code helps Mezzacapps's team study the magnetic field in Type II supernova explosions (Figure 3(b)) as well the broader scientific community.
- Specification of a seed point list from a text file is now supported.
- Paul Sutter (UIUC) reported that FLASH data was not being processed correctly. It turns out that FLASH has changed their units and the data they are producing is at a scale that doesn't work on most graphics cards. We modified VisIt to have just-in-time scaling code so that the data is scaled down as it is rendered. This fix was rolled out in VisIt 2.1.
- Movie showing particle advection through magnetic field of seed points randomly distributed on the supernova shock surface.

2.3 High Energy Physics – Accelerators

The Problem. A recurring theme through many science projects is the growth of data size and complexity outpacing the capabilities of important and familiar analysis toolsets.

The Solution. VACET has a long-term mission objective of providing production-quality, petascale capable visual data analysis infrastructure to DOE's HEP community, in particular the SciDAC 2 project entitled Community Petascale Project for Accelerator Science and Simulation (COMPASS). In the past reporting periods, we have documented progress towards this objective with several different teams within COMPASS. This time, we are reporting new work with a subset of COMPASS at Fermi lab, specifically, Jim Amundson, Eric Sterne, and Panagiotis Spentzouris who are workng with the Synergia code. Our efforts are aimed at helping this team use VisIt to perform visual data analysis of the data produced by Synergia.

The Impact. Our work is providing the Fermi team with new capabilities that will enable them to perform visual data analysis of data being produced by the Synergia code. This objective is not possible for them using legacy tools.

Accomplishments.

Hank visited Fermi Lab in early October 2010. He spent a day doing give and take where we tried to make movies, do exploration, debugging, etc.

This trip showed that three capabilities are critical to their work flow. They are:

- Data binning: being able to construct objects like 1D, 2D, and 3D histograms, PDFs, and averages.
- Selections: being able to identify a set of particles (note: Brad Whitlock implemented these on behalf of Allen Sanderson for a fusion project, hence his inclusion as a virtual member on this project).
- "Persistent Particles:" being able to watch the locations traveled by a set of particles.

Note that all three of these features have been developed by VACET personnel. Data binning and selection had been significantly improved already for VisIt 2.1, the timing of these new features and capabilities was fortuitous for this visit to the Fermi team.

The Q&A session revealed some critical bugs:

- The data binning operator did not work with the volume or contour plots.

- Defaults for data binning could not be saved into session files.
- The persistent particles operator only worked in conjunction with selections when dealing with FastBit data.

Hank fixed these three issues. The fix will be available to Fermi users in November, when VisIt 2.1.1 is released.

2.4 Groundwater

The Problem. The PFLOTRAN code models fluid flow through porous media. Its primary developers are Peter Lichtner and Glenn Hammond. This code is widely used in the EM community for computing the movement of fluids, especially fluid transport of contaminants. Other PFLOTRAN stakeholders come from the DOE-EM funded project entitled Advanced Simulation Capability for Environmental Management ASCEM (see Section 2.5). The "problem" here is that supporting any community of users requires ongoing support and consulting to ensure the community's specific needs are met.

The Solution. Work has been proceeding on various fronts. Initially, our goal was to generate a data file reader capable of supporting their data model and format with a scalable implementation. This effort has been successful, and includes code contributions to the data reader back to the VisIt project from ASCEM visualization developers. Ongoing work supports the team with general help, as well as implementing features specific to their usage paradigms.

The Impact. In the past, the PFLOTRAN code team and many of its users employed other visualization tools like Tecplot. We have provided them for their day to day needs a more fully featured, scalable tool for their analysis and visualization, with room for growth (e.g. as they proceed to AMR-based models).

Accomplishments.

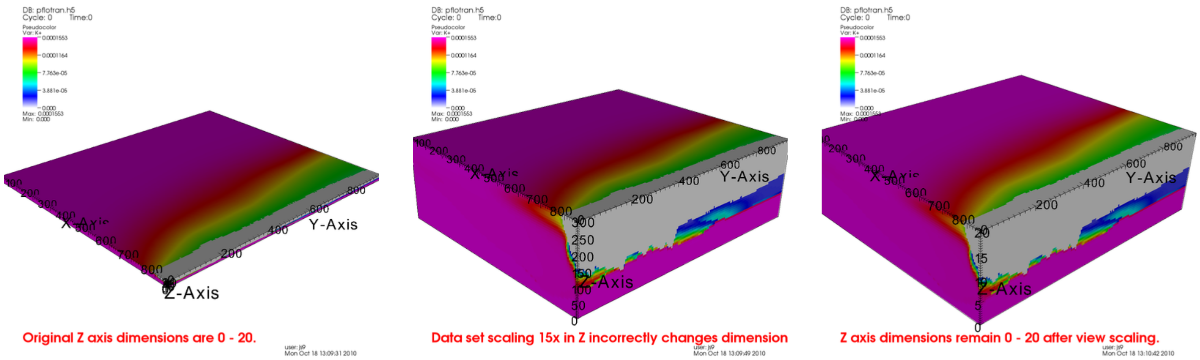
2.5 Environmental Management

The Problem. DOE's EM-funded project entitled Advanced Simulation Capability for Environmental Management (ASCEM) was formed to develop transformational, high performance computer modeling capabilities to better meet the challenge of waste disposal and cleanup left over from the creation of the US nuclear stockpile decades ago. There is a strong visual data analysis and exploration component to ASCEM; ASCEM has adopted VisIt for project-wide use.

The Solution. As with other user groups, achieving effective community-wde adoption requires a combination of domain-specific feature enhancements, consulting in day-to-day and advanced use, and assistance with the creation of high-end movies and images. In the short term, our primary efforts thus far have been consulting and in integrating VisIt PFLOTRAN loader code contributions from the ASCEM visualization team.

The Impact. By adopting VisIt for project-wide use, ASCEM can be assured that they have production-quality, visual data exploration and analysis software that will scale to very large data sizes, and that can accommodate advanced analysis capabilities like support for multivariate statistics and analysis across ensemble collections of data. We contributed to the generation of the ASCEM Phase I report, which describes accomplishments during the first year of ASCEM operation.

Accomplishments. Our primary work this period is providing support to the ASCEM visualization team. There are two primary results to date. The first is visual data exploration of historical observational data (Figure 5). The other is in using VisIt to explore uncertainty across an ensemble collection of data generated by Monte Carlo simulation runs (Figure 6). In both these



(a) The original data set prior to VisIt bug fixes. Note that for this type of data (groundwater simulation), the height axis is much shorter than the horizontal axes and must be expanded for effetive visual analysis. This problem arises in other domains, such as atmospheric and climate simulations.)

(b) An illustration of a "traditional" solution: scale the data by 15x in the Z dimension, for example. Visually, the immediate problem is that the Z axis is now labeled incorrectly. Further problems exist, though: any analysis algorithms performed on this data set will be affected by the rescaled dimension, and will generally result in incorrect answers. Data set scaling is thus not an appropriate solution.

(c) An illustration of the new solution implemented in VisIt: the data set is scaled only visually. Rendering, lighting, and interactions use the scaled dimensions, while axis labeling uses the original dimensions. And most importantly, all numerically based analysis use the original dimensions and therefore remain correct, as if no scaling had occurred.

Figure 4: Before and after images illustrating the problems resulting from vastly different sizes in data domains and our solution.

cases, VACET provided consulting support to the ASCEM team, and both of those images were produced by the ASCEM visualization team.

2.6 Fusion: Particle Selection and Analysis

Background: VACET has a long-standing mission objective of helping to meet visual data analysis and exploration needs of the fusion science community. The work here reports on progress on work with our fusion science collaborators.

Fusion scientists are running simulations that currently use millions to billions of particles with multivariate values at each particle. These scientists would like to have the ability to explore the nature of the particle orbits in an interactive manner. However, it is impractical to view this many particles at once and gain any meaningful information. The fusion researchers' desire is for software tools that will allow them to isolate and study "interesting" particles, where interesting may be defined with a compound boolean data or spatial range query or using statistical methods. In addition to the particles themselves, fusion scientists would like the ability to display other non-particle data concurrently with the particle-based data. Such non-particle data could be scalar (electric potential) or vector (magnetic field).

The Problem. S. Either and Ku ***What is this person's name? What institution?*** they are aiming to understand the mechanisms for radial transport in Gyrokinetic simulation codes. It is well known that particles become trapped within electromagnetic waves and are transported radially. However, not all particles exhibit such a behavior (of becoming trapped and untrapped) yet still have radial transport.

The Solution. Parallel-capable visual data analysis and exploration software that provides

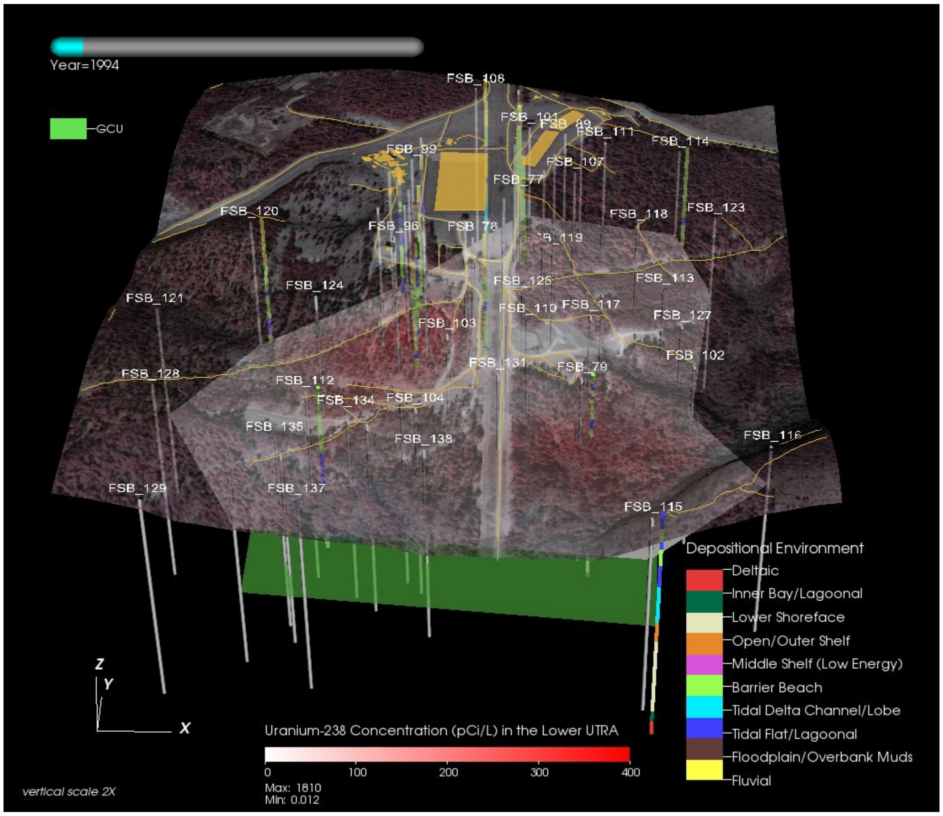


Figure 5: Historical field data from the F-area seepage basin at SRNL. This image shows a diverse collection of data: observational wells that are color-coded by depositional environment, U238 concentration within one of three different aquifers, the location of one of three different aquifers (green), satellite imagery layered onto surface topography, GIS data (roads, buildings). This image was created by J. Horsman (LBNL), an ASCEM visualization developer, using VisIt.

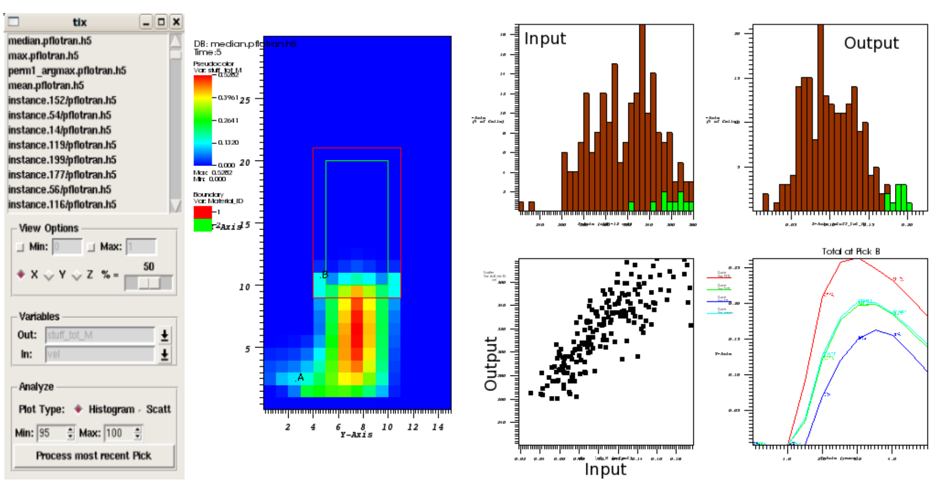


Figure 6: Plots generated from prototype UQ visualization GUI: concentration slice, input and output histograms, scatter plot, and central tendencies and selected quantiles of time varying concentration at a selected location. In this example, a user-written Python script produces the GUI on the left, and then directs VisIt via the Python interface to produce the five plots/visualizations.

the means for a scientist to quickly formulate compound boolean range queries to quickly locate interesting particles in large datasets, and to perform more in-depth exploration and analysis on that interesting subset. The interface for specifying queries is a parallel coordinates implementation that can accommodate very large datasets[4]. Finding and managing this selected subset that spans multiple timesteps is the result of a related VACET project (Section 4.1) that is brought to bear on specific fusion science problems here.

The Impact. Previously, it has been challenging for fusion researchers to understand particle behavior. They can take measurements and plot particle paths, but even then, it is difficult to gain insight. Our fusion stakeholders are aiming to use our new query-based capabilities to do more precise and detailed analysis. Early results from these stakeholders are shown later in Section 4.1, Figure 17.

2.7 Fusion – Poincaré Topology Analysis

Physicists are currently studying the affects of magnetic islands that form in burnding plasma by studying results of numerical simulations of magnetically confined fusion. These islands cause defects in the magnetic field and the current flow resulting in contact between previously separate regions. This contact results in "hot areas" coming into contact with "cool areas," which leads to core cooling. Physicists would like to have tools that allow them to be able to automatically generate Poincaré maps of the magnetic field and detect formation of magnetic islands and to track/analyze them over time.

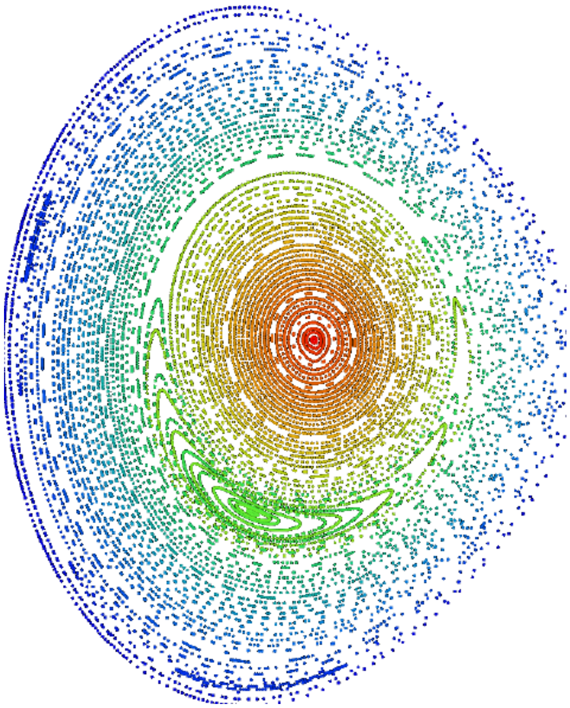
To further aid in studying this phenomena, physicists would like to be able to see other types of data displayed concurrently with the Poincaré maps. This data may be scalar (electric potentials) or vector data (magnetic fieldlines) and may have its own visualization requirements. For instance, the scalar data may be viewed using a variety of techniques, such as volume rendering, slicing or isosurfaces.

The Problem. The process of generating Poincaré plots requires using a dense set of points to form a continuous representation of the cross section of a magnetic surface. The large number of points is bound by computational resources as well as numerical accuracies (integration and interpolation). Further, these plots currently offer no topological analysis. As such, any analysis is manual in a non-interactive setting. Previous approaches, which involve creating "puncture plots" and then a process of manual inspection, do not provide the desired level of accuracy nor do they provide the ability to perform quantitative analysis of magnetic island formation over time. Figure 7 shows the traditional puncture plot along with results produced by our new method.

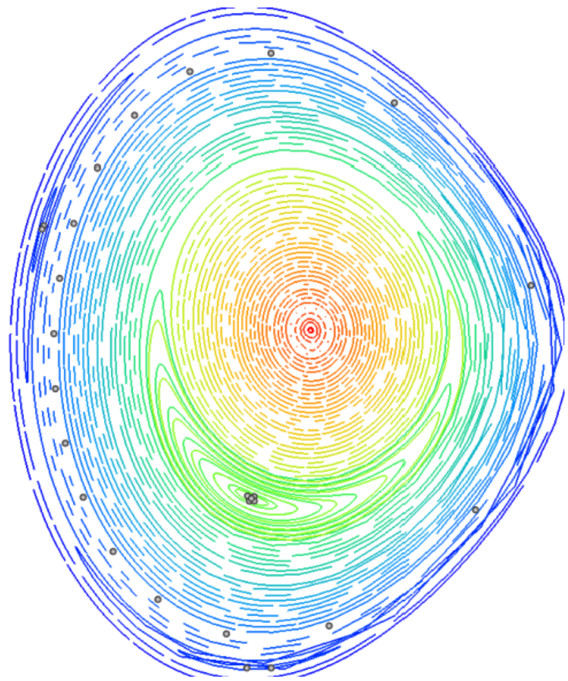
The Solution. In a collaborative effort with fusion scientists, our team has developed new tools that enable interactive analysis of magnetic islands. The new algorithm computes a continuous representation using a near minimal set of puncture points to reduce numerical inaccuracies while providing topological analysis. The new technology is delivered in petascale-capable, production-quality form in the VisIt visual data analysis and exploration application to the DOE and worldwide science communities.

The Impact. Our new method helps increase scientific understanding by conveying fine structure, which is necessary for understanding heat transport in highly stochastic fields, along with the ability to compute and analyze topological features like magnetic islands. Previous approaches that use "puncture plots" produce images where results were inconclusive and ambiguous. lize islands within islands. Not only does our technique visualize them but is also able to identify them. Seeing such fine structures is important for understanding heat transport in highly stochastic fields.

[4]The parallel coordinates technology for very large data was the subject of a Supercomputing 2008 paper from VACET researchers.



(a) The traditional puncture plot produces visual results that are difficult to interpret and offer no means for quantitative analysis.



(b) Our new method computes the locations of magnetic islands and provides the ability to perform quantitative analysis of these features over time.

Figure 7: Comparision of traditional puncture plots and our new method. This type of capability helps fusion researchers to gain insights into complex data produced by simulations.

Accomplishments.

The major accomplishments for this period have been in the continued refinement of the underlying numerical methods we use for generating the fieldlines (e.g., streamlines and integral curves), which are input to the Poincaré analysis.

Another accomplishment has been identification of another key periodic topological component, which we are currently incorporating into the analysis code.

This work has been accepted for presentation and publication as part of IEEE Visualization and the APS-DPS meetings.

2.8 Nuclear Science

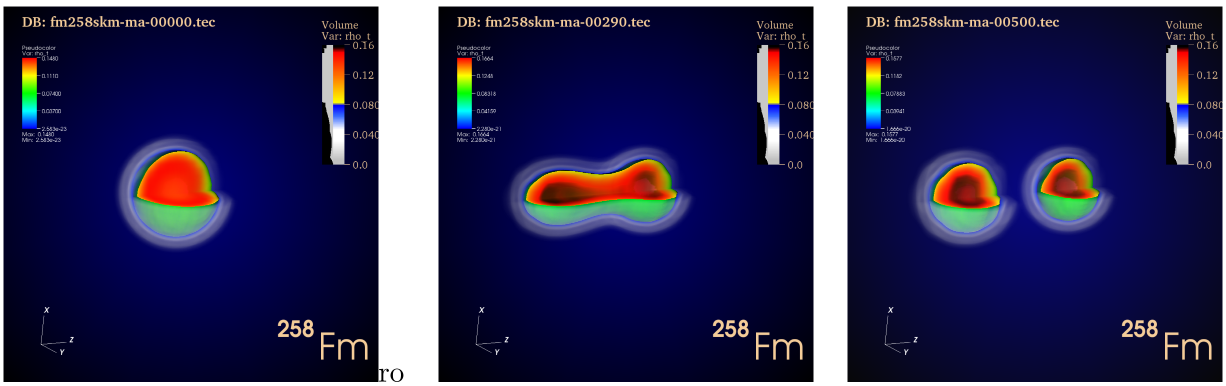
The Problem. There are approximately 3,000 known nuclei, most of them produced in the laboratory. It is estimated that additionally up to approximately 6,000 nuclei could in principle still be created and studied in the foreseeable future. An understanding of the properties of these elements is crucial for a complete nuclear theory, for element formation, for properties of stars, and for present and future energy and defense applications.

The Solution. We are providing one-on-one consulting to researchers from the University Nuclear Energy Density Functional project in the use of VisIt as well as in applying principles of visualization to explore complex data. For example, we are providing help to one UNEDF researcher in the area of multivariate visualization of energy density functional parameter relationships. The goal is to find new relationships that will lead to improvements in the energy density functional and its nuclear shape deformation predictions.

The Impact. While this collaboration is in its early stages, the hope is that improved visual data exploration and analysis capabilities will help to foster Discovering new relationships may that lead to better understanding of energy density functional corrections and better nuclear shape deformation predictions.

Acomplishments.

- Worked with a member of the UNEDF team to introduce him to VisIt, and helped him setup some visualizations of his data. He was able to produce several movies, a few frames of which are shown below (Figure 8).



ro

Figure 8: These are three images from an animation created by an UNEDF researcher.

- Small Multiples: It was discovered that some infinite matter parameters are related to the surface parameters. We are currently following up with Mario Stoitsov on the consequences to the energy density functional. We used EVEREST, a high-resolution tiled display, to simultaneously display many of these pairwise displays (Figure 9).

3 Technology Incubation Projects

3.1 Flow Visualization: Streamlines, Stream Surfaces, Pathlines

The Problem. Modern simulation codes increasingly involve the simulation of vector fields that appear in problems relating to astrophysics, thermal hydraulics, fusion, and fluid flow. Existing visualization and analysis capabilities for scalar fields cannot fulfill the need for adequate visualization of such vector fields, as the delocalized nature of typical phenomena described by vector fields such as e.g. transport and mixing is not adequately captured by scalar-based approaches. For such problems, modern integration-based methods can be used to derive visualization and analysis from the trajectories of idealized massless particles. These methods provide much increased insight into vector fields, but are computationally intensive and challenging in their application to very large (petascale) datasets. The aim of this project is to enable physicists, chemists and fluid dynamicists to visualize and analyze their state-of-the-art simulation data using robust, efficient and scalable integration-based visualization techniques over a wide variety of data representations and problem characteristics.

The Solution. Our solution is based on a novel code framework, integrated into VisIt, that enables the efficient and scalable computation of integral curves in vector fields represented over regular, structured, unstructured and AMR meshes. Due to the large variation in problem characteristics involving integration-based visualization algorithms, our framework can leverage several

Figure 9: EVEREST display of over 100 pairwise parameter relationships with fitted nonparametric spline and shaded uncertainty intervals produced with R graphics.

distinct parallelization schemes that are uniquely suited to specific problem parameters. Also, we provide an adaptive scheme that allows scientists to conduct vector field visualization without having to familiarize themselves with parallelization strategies, allowing for rapid adoption of integration-based methods.

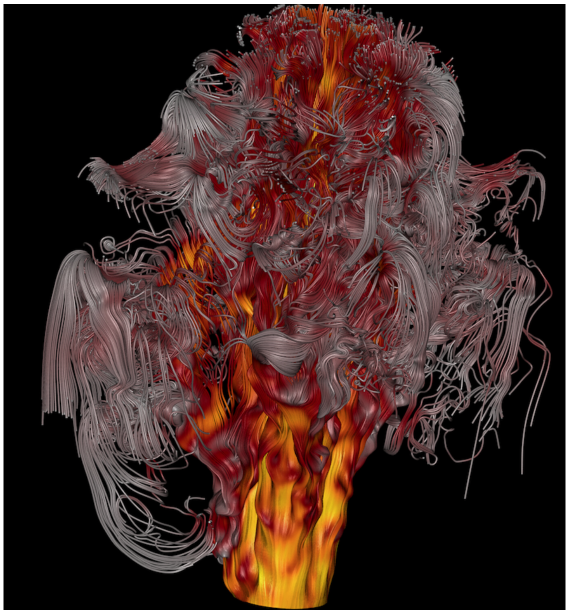
To increase performance and scalable efficiency, two of the parallelization schemes make use of hybrid parallelism, where fewer MPI tasks are used (typically one per node) and shared-memory parallelism is employed within a node. This approach, although more challenging to implement, can enable significant performance and efficiency gains on supercomputers based multi-core CPUs.

Furthermore, building on these capabilities, we are investigating novel integration-based visualization algorithms, such as integral surfaces and Lagrangian techniques, that can provide adequate abstraction for specific vector field phenomena. The aim of this project to deliver a comprehensive solution for integration-based visualization based on both baseline integration capabilities as well as advanced visualization algorithms to science stakeholders in the VisIt visualization tool.

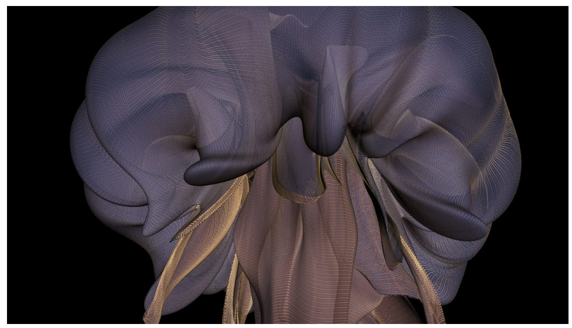
The Impact. Based on our framework, scientists will be able to leverage modern integration-based visualization techniques to visualize and analyze vector field data to investigate phenomena such as transport and mixing. Our work specifically addresses very large / petascale data, to enable robust analysis on current and future-generation datasets. Furthermore, we will provide necessary infrastructure and tools to build improved visualization tools that address the need to analyze simulations of ever-growing complexity.

The hybrid parallelism implementation results in much improved performance, requires less communication and I/O bandwidth than a traditional MPI-only implementation. We have speedups ranging from two to ten time that of the MPI-only implementation. This will allow better use of supercomputers time, it will allow large numbers of streamlines to be used for better data understanding and the use of larger dataset sizes to be tested. Figure 10 show example output from this improved algorithm.

Moreover, recent software re-engineering efforts have resulted in an improved code base that is much more flexible and efficient than its previous incarnation. This framework is integrated with VisIt and thus customers benefit directly from these improvements.



(a) 10,000 streamlines in a high-speed jet flow demonstrate the capabilities of the improved integration framework deployed in VisIt 2.1. The image is computed an rendered in seconds on a commodity workstation.



(b) A high-resolution streak surface computed using 160,000 individual particles. This image is a single frame of an animation that won a People's Choice Award at SciDAC 2010's Visualization Night.

Figure 10: The new parallel-capable integral curve calculation code in VisIt offers the ability to process very large datasets on parallel machines.

Accomplishments.

- Improved handling of AMR data and unstructured meshes, deployed in VisIt 2.1.
- General flexibility improvements and large efficiency gains in integration framework, deployed in VisIt 2.1.
- Basic research on illustrative rendering of integral surfaces to improve visualization and obtain additional insight.
- Basic research on treating integration in unstructured meshes on GPUs (integration into VisIt is considered but not currently a specific goal).
- Many of the "rough edges" that users stumble upon were made "foolproof.". This capability will be appear in production form in VisIt 2.2.

Flexibility. The previous incarnation of this code was basically exclusively dedicated to streamlines and Poincaré analysis. Through the efforts of Hank Childs, Christoph Garth, Dave Pugmire, and David Camp, this code has transformed in the last six months into arbitrary particle advection code. Previously, the entire path of each streamline was recorded, along with all state that might be needed for rendering. Now, the code is "lean and mean:" it is possible to advect particles and

do any custom analysis you please, including analysis that only stores a few bytes per particle (regardless of the number of steps). Six months ago, the key code resided in avtStreamlineFilter. Now, we have an avtPICSFilters (PICS = parallel integral curve system), with avtStreamlineFilter being a derived type. Each time a particle is advected, a virtual function is called to perform analysis for that step. For streamlines, the new location is recorded, as well as information needed for rendering. For Poincaré analysis, the step path is checked against the location of a plane. Previously, these routines were conflated with a "bloated" result. The refactoring, which took months of man-effort, is a result of the recognition that particle advection is a field that goes considerably beyond simply plotting streamlines. It is also in recognition of the need to make sure that evaluations, interpolations, calculations, and storage are "lean and mean." The new system is flexible and powerful: there is a choice of integrator (Dormand-Prince or Adams-Bashforth), a choice of seed locations, a choice of termination criteria, a choice of parallelization, and a choice of how to analyze the particles path. This work helps to pave the way for enabling advanced analysis of flow fields, such as our work with Fusion scientists (Section 2.7) as well as new visualization techniques for flow field data (Figure 11).

Performance. For each step, the integrator must do multiple evaluations of points in space, which includes both cell location and interpolation. Cell locations are assisted by data structures and an important question is how much time should be spent preparing this data structure. The more lookups that are performed, the more the data structure initialization time can be amortized over those lookups. As it turns out, streamline evaluation performs enough lookups that it is beneficial to construct data structure that is heavier weight than the standard "oct-tree" scheme. In his IEEE Visualization 2010 paper, Christoph Garth introduces a new scheme for lookups that, despite having larger initialization time, is so fast that streamline performance overall is faster. For Poincaré analysis, this improved lookup time is especially important, since the particles are advected for such a long period of time. Allen Sanderson reports that the Poincaré analysis he performs frequently dropped from 20 minutes to 25 seconds when using this new scheme.

Foolproofing. VisIt's streamline code has been characterized as "buggy." But we believe, although there were previously many bugs, the real problem is that users must understand how to navigate its many controls as an expert. The user needed to understand step lengths, tolerances, distances to integrate, etc. If any of these were incorrectly set, the result could be an integration that would require millions of steps, which would likely exhaust memory. To fix this problem, we changed the controls so that step lengths, tolerances, etc, are all proportional to the data size. Further, we changed the termination criteria to prevent "runaway" cases; VisIt will now only integrate millions of steps if explicit told to by the user. All of these changes will be available in VisIt 2.2.

3.2 Hybrid Parallelism and Volume Rendering at Extreme Concurrency

The Problem. Modern computational platforms are evolving towards using multi-core processors; future generations of machines will be built using processors containing tens or hundreds of cores. There is concern that existing, traditional message-based parallel programming models will not scale well on such platforms. The aim here is to better understand how well hybrid-parallelism, which combines both traditional message-based distributed memory parallel concepts with multi-core, shared-memory parallelism, performs for visualization algorithms, raycasting volume rendering specifically, as compared to traditional message-based, distributed-memory parallelism.

The Solution. Our solution entails conducting performance and scalability tests of traditional and hybrid parallel implementations of raycasting volume rendering, a staple visualization algorithm. Our approach is to compare performance using several different metrics: (1) abso-

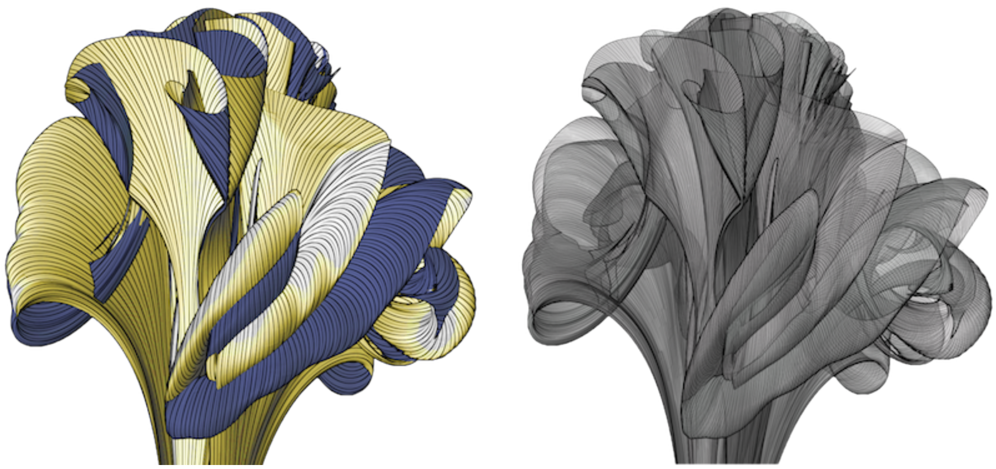


Figure 11: Illustrative techniques are employed to render integral surfaces – in this case, a path surface. Adaptive-resolution surface texturing is used to indicate the flow direction on the surface.

lute runtime, (2) memory footprint at various stages of algorithm execution; (3) communication characteristics of the two implementations.

The Impact. The results of this study show that the hybrid-parallel approach offers clear and distinct performance advantages when compared to the traditional approach to parallelism. First, the hybrid-parallel version consumes between one sixth and one twelfth the amount of memory required by the traditional MPI version just for initialization. Second, at high levels of concurrency, the hybrid-parallel implementation runs about three times faster. Third, the hybrid-parallel version requires only about half the communication bandwidth compared to the MPI version. These early results will help to shape the architecture of future visualization and analysis applications so as to be able to run effectively on current petascale and future exascale platforms: the hybrid-parallel approach allows for tackling larger problems on a given set of resources than is possible using an MPI-only approach.

Accomplishments.

- Strong scaling study. A strong-scaling study aims to discover how time-to-solution improves as more and more processors are brought to bear on a fixed-size problem. The expectation is that as you bring more processors to bear on a problem, time-to-solution should decrease. The focus of our strong-scaling study was to discover how the relative time-to-solution decrease differs between MPI-only and MPI-hybrid approaches over a range of concurrencies.

  In our tests, we used a fixed size dataset of 4608^3 over concurrency levels ranging from 1728 to 216,000-way parallel, the highest level of concurrency ever published. These tests, run on jaguarpf at ORNL, reveal that: (1) for just the initialization phase and before we begin any actual volume rendering work, the MPI-only implementation consumes 12x more memory than the MPI-hybrid implementation due to MPI overhead; (2) the MPI-hybrid implementation requires 40% less memory for ghost data at runtime owing to use of larger-sized data blocks; (3) the MPI-only implementation sends about 6x more communication messages during the compositing phase as compared to the MPI-hybrid implementation; (4) at the highest level of concurrency, the MPI-hybrid implementation is about 3x faster than the MPI-only implementation due primarily to the reduced communication load. Figure 12 shows visual output from a 216K-way parallel run.

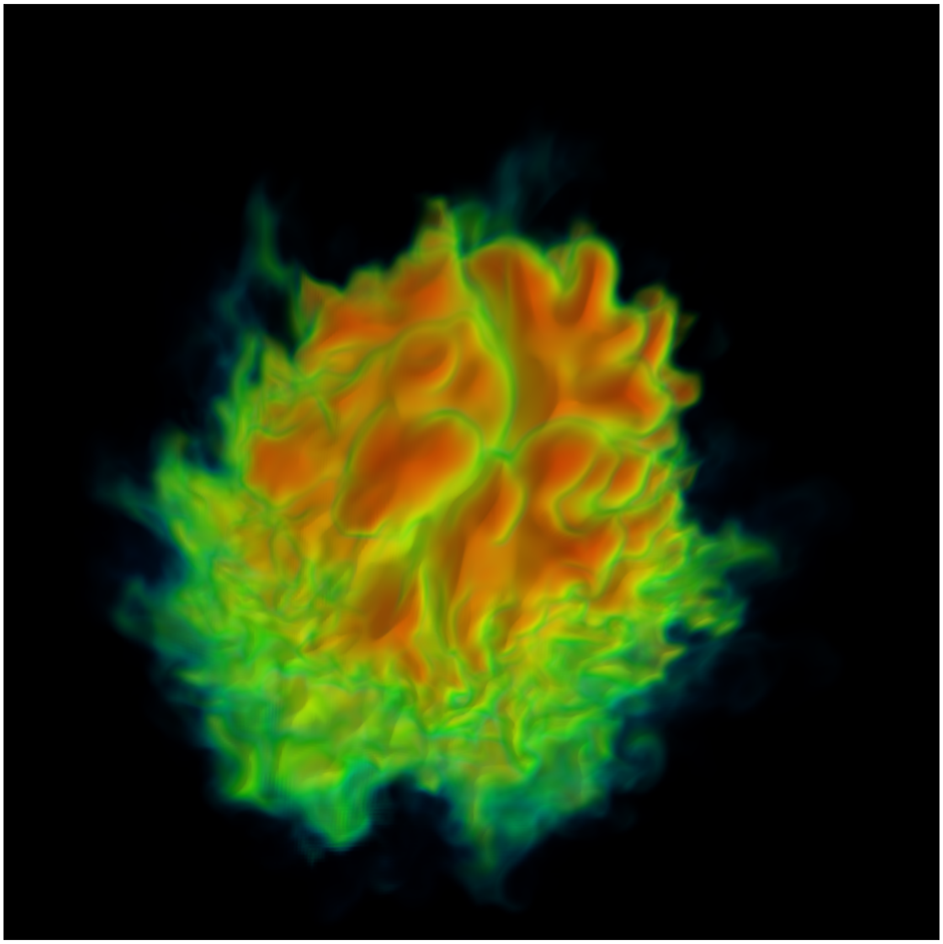


Figure 12: Our hybrid-parallel volume rendering application produced this image of combustion simulation data using 216,000 cores on JaguarPF.

- Weak scaling study. In contrast to the strong-scaling study, the weak-scaling study aims to increase problem size while adding more processors. The idea is that using more processors, and their associated memory, enables tackling ever-larger problems, which is important from a large-data perspective since we wish to use large, parallel computational resources to enable visual data exploration and analysis. Ideally, time-to-solution remains constant at increasing concurrency for a code that exhibits perfect weak scaling characteristics.

  In our weak scaling studies, we use a 384^3 data block per processor at all concurrency levels. At 1728-way parallel, the resulting mesh is 4608^3, and at 216,000-way parallel, the mesh is 23040^3 in size (12.2 trillion cells). We observe that this mesh resolution far exceeds that of current scientific codes.

  With raycasting volume rendering, there are actually two dimensions of "problem size." The first is the size of the source data itself. The second is the size of the resulting image. In order to better understand weak-scaling characteristics of our hybrid-parallel application, we explored two types of weak scaling: weak-dataset scaling and weak scaling. With weak-dataset scaling, we use a fixed-size image of 4608^2 pixels at all levels of concurrency. With weak scaling, we also increase the resolution of the image from 4608^2 pixels at 1728-way parallel up to 23040^2 pixels at 216,000-way parallel.

The results of the weak scaling studies confirm the favorable MPI-hybrid characteristics first observed in the strong-scaling studies: (1) at initialization, the MPI-hybrid implementation requires 12x less memory than the MPI-only implementation due to MPI overhead; (2) during the raycasting phase of the algorithm, the MPI-hybrid approach shows better scalability than the MPI-only approach; (3) during compositing, the MPI-hybrid approach requires substantially less communication traffic than the MPI-only approach; (4) the MPI-hybrid implementation consistently runs faster than MPI-only, as much as about 40% faster at 216,000-way parallel on the 12.2 trillion-cell dataset.

- Many-core GPU study. In an effort to better understand how these performance gains would extend to architectures having more than a few cores per processor, we conducted a test run on a multi-GPU system (Longhorn at TACC). Our testing configuration consisted of running an MPI-hybrid implementation in parallel across 448 GPUs, each of which is capable of concurrently executing 30 CUDA thread blocks, producing an effective concurrency of 13,440-way parallel. The main objective here is to evaluate the relative performance of our MPI-hybrid implementation run a distributed memory GPU cluster as compared to a traditional distributed memory multi-core CPU system. The results indicate that the 448-GPU configuration runs about 30% faster than on a distributed memory multi-core system at similar concurrency.

3.3 Uncertainty Visualization

The graphical depiction of uncertainty information is emerging as a problem of great importance. Scientific data sets are not considered complete without indications of error, accuracy, or levels of confidence. The visual portrayal of this information is a challenging task. This work takes inspiration from graphical data analysis to create visual representations that show not only the data value, but also important characteristics of the data including uncertainty. The canonical box plot is reexamined and a new hybrid summary plot is presented that incorporates a collection of descriptive statistics to highlight salient features of the data. Additionally, we present an extension of the summary plot to two dimensional distributions. Finally, a use-case of these new plots is presented, demonstrating their ability to present high-level overviews as well as detailed insight into the salient features of the underlying data distribution.

The Problem. As data becomes increasingly large and complex, visualization and data analysis techniques are required that not only address issues of large scale data, but also allow scientists to better understand the processes that produce the data and the nuances of the resulting data sets. Uncertainty, in the form of confidence, variability, and error, as well as model bias and trends, is regularly included within data sets and is used to express descriptive, qualitative characteristics of the data. Because uncertainty is crucial in understanding the reliability of information and thus in objectives such as decision making, its absence can lead to misrepresentations and incorrect conclusions. Too often, traditional visualization approaches overlook available uncertainty information. As the importance of visualizing these large, complex data sets grows, the actual task of visualizing them becomes more complicated; incorporating the additional data parameter of uncertainty into the visualizations becomes even less straightforward. Difficulties in applying preexisting methods, additional visual clutter, and the lack of obvious visualization techniques leave uncertainty visualization an unsolved problem.

The Solution. The goal of this work is to create a summary plot that incorporates higher order descriptive statistics to concisely present data with uncertainty information. This work takes inspiration from the visual devices used in exploratory data analysis and extends their application

to uncertainty visualization. The statistical measures often used to describe uncertainty are similar to measures conveyed in graphical devices such as the histogram and box plot. This research investigates the creation of the summary plot, which combines the box plot, histogram, a plot of the central moments (mean, standard deviation, etc.), and distribution fitting. The box plot has a canonical feel; the "signature" of the plot is easily recognizable and does not need much explanation to allow for a full understanding. The focus of this work is to create a summary plot that similarly incorporates higher-order information, allowing for the quick identification of characteristic features. This higher-order signature provides at-a-glance recognition of variations from normal and allows easy comparison of data distributions in detail. In addition, a 2D extension of the summary plot is presented, which provides for the comparison of correlated data.

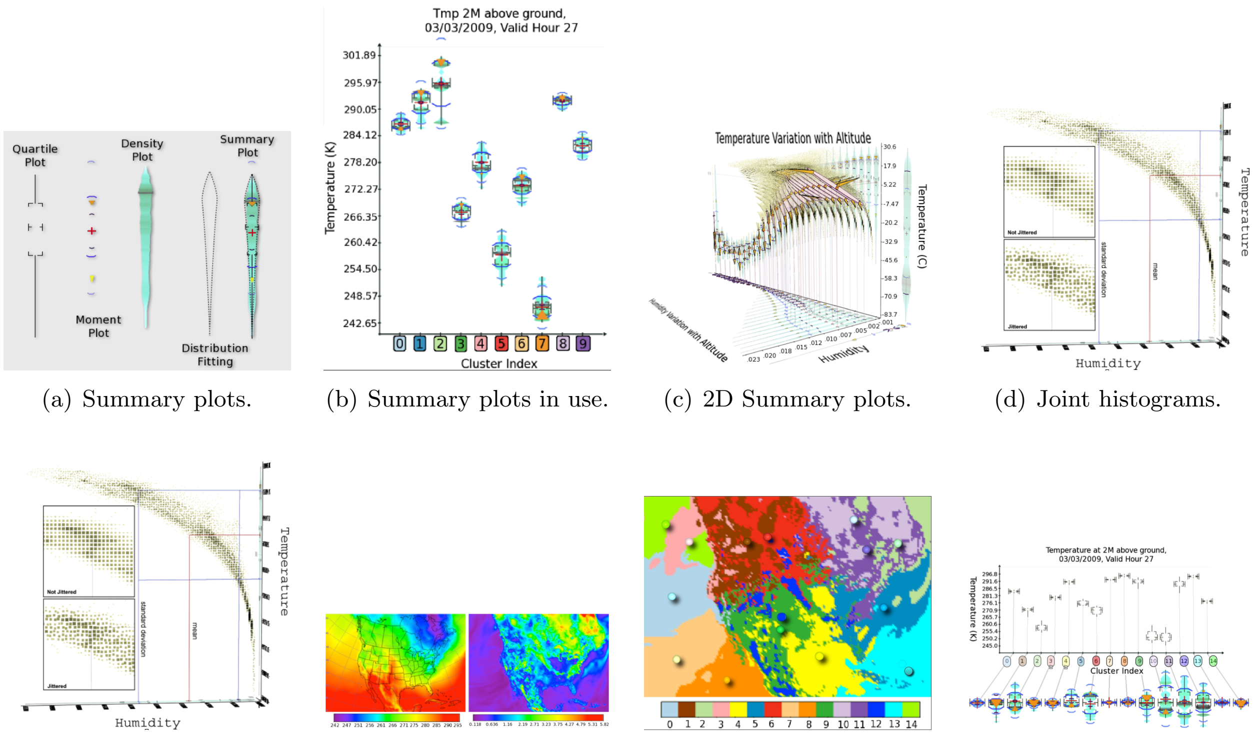
The Impact. Uncertainty information has been inadequately addressed in the visualization community, largely because of the difficulties involved with visually expressing this additional data. If visualization is to become a robust decision-making tool, it must represent uncertainty, in some form, to the audience. This work provides a method for investigating visual characteristics of a data distribution, both for learning about the shape of the data set and for expressing the associated uncertainty.

Project Summary. The hybrid box plot we are introducing can be more formally titled the summary plot. This display includes not only the quartile information present in the form of a modified box plot, but also a collection of descriptive statistics and density information. We use an abbreviated form of the traditional box plot to convey the 5-number summary and a symmetrically drawn, colormapped, histogram to show density information. Descriptive statistics, in the form of mean, standard deviation, and higher-order moments, are expressed as glyphs with the design of each reflecting the semantic meaning of the statistic. Finally, distribution fitting capabilities are added to allow the user to compare against and find a best fit from a library of well-known distributions.

2D Summary Plot. In addition to a statistical summary for a 1D categorical data set, users require methods for comparing multiple, correlated data sets to understand how data values are related. Here, we explore methods for summarizing categorical data with pairs of values associated with each sample. The joint summary places 1D summary plots for each data set perpendicularly to orient the viewer. Joint mean and standard deviations, a joint histogram, and a reduced higher-order moment plot are added, providing a display which shows the relationship between correlated data sets. Figure 13 shows representative output from this method.

3.4 Topology: Edge Maps

Robust analysis of vector fields has been established as an important tool for deriving insights from the complex systems these fields model. Many analysis techniques rely on computing flow trajectories, a task often hampered by numerical instabilities. Approaches that ignore the resulting errors can lead to inconsistencies that may produce unreliable visualizations and ultimately prevent in-depth analysis. We propose a new representation for vector fields on surfaces that replaces numerical integration through triangles with linear maps between its edges. This representation, called edge maps, is equivalent to computing all possible streamlines at a user defined error threshold. In spite of this error, all the streamlines computed using edge maps will be pairwise disjoint. Furthermore, our representation stores the error explicitly, and thus can be used to produce more informative visualizations. Given a piecewise-linear interpolated vector field, there are only 23 possible map classes for a triangle, permitting a concise description of the flow. This work describes the details of computing edge maps, provides techniques to quantify and refine edge map error, and gives qualitative and visual comparisons to more traditional techniques.



(a) Summary plots. (b) Summary plots in use. (c) 2D Summary plots. (d) Joint histograms.

(e) Co- and skew-variance. (f) Mean (left) and standard deviation of temperature. (g) Clusters based on variance. (h) Summary plots of cluster locations.

Figure 13: The Summary Plot and its use.

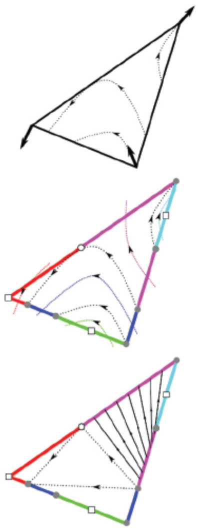
The Problem. Vector fields are a common form of simulation data appearing in a wide variety of applications ranging from computational fluid dynamics (CFD) and weather prediction to engineering design. Visualizing and analyzing the flow behavior of these fields can help provide critical insights into simulated physical processes. However, achieving the consistent and rigorous interpretations of vector fields is difficult, in part because traditional numeric techniques for integration do not preserve the expected invariants of vector fields.

The Solution. We have devised a new data structure, edge maps, for storing the flow behavior of a vector field that does not rely on numerical integration. This structure is complementary to the traditional way of storing vector fields as piecewise linear interpolations over a mesh. Each triangle stores a map that encodes the inflow/outflow behavior over the boundary of the triangle. This allows us to replace the notion of integration with a different primitive, map lookup.

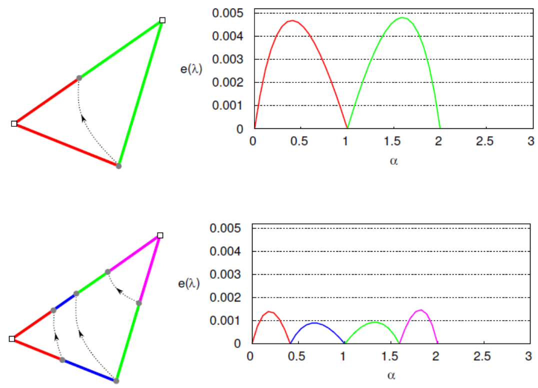
The Impact. Our contributions include (1) The definition of edge maps for triangulated 2D vector fields, and an algorithm to compute the approximate edge maps; (2) Quantification of error bounds with this approximation; (3) A refinement procedure for reducing mapping error; and (4) New visualizations of flow instabilities using edge maps.

Accomplishments.

This work is a collaboration with the DOE/ASCR MAPD (Mathematics for Analysis of Petascale Data) project joint with Sandia National Labs, Texas A&M, and the University of Utah. While the basic research is being claimed under MAPD, VACET will be concerned with the deployment of new technologies produced by the effort. We anticipate the research will lead to new strategies, algorithms, and software technologies for handling petascale data by applying topological techniques that synthesize approaches for handling uncertainty.



(a) Discrete flow map for a triangle. (top) The original triangle is represented as three vectors, which impose a flow throughout the interior. (middle) Our representation subdivides the boundary into a set of intervals. Intervals are broken at internal transition points (white circles), external transition points (white squares), and additional points (grey circles). (bottom) Pairs of intervals are group into links which represent sources and destinations of flow through the triangle. Flow paths can be approximated through each link by linearly interpolating position from the source interval to the destination.



(b) Top row: The red and green error curves show the mapping error in red and green links of the triangle. The mapping error is drawn as a function of arc-length parameter of the triangle, α, counterclockwise from the bottom left vertex. For $2.0 < \alpha < 3.0$, there is no mapping error, since this segment of the triangle is acting as a destination. The average edge-length of the triangle in consideration is 0.025. Bottom row: A refined map contains much smaller mapping error when refined to $\delta = 0.0025$ as compared to the basic map. During refinement, each link gets split into two links each.

Figure 14: Edge maps and error measurement/mapping.

Mathematical Foundations developed regarding flow maps (Figure 14(a)). We performed an in-depth analysis of the possible flow behavior using piecewise linear flow across a single triangle, and evaluated the mapping error incurred by this representation.

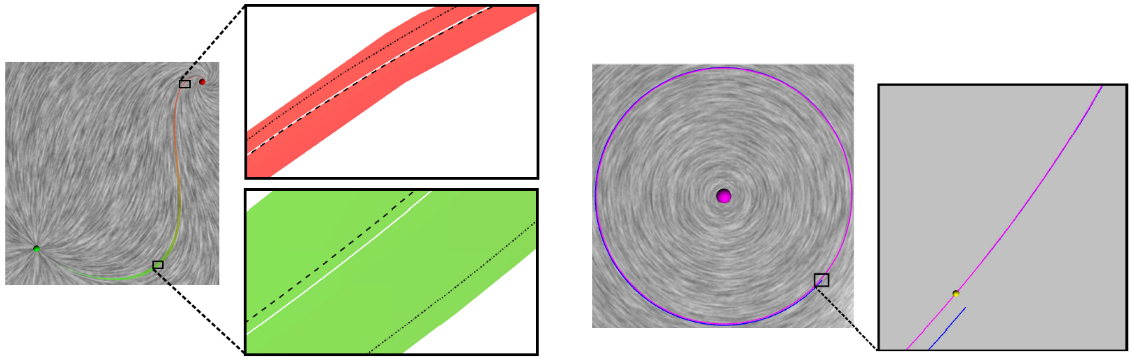
Map Refinement (Figure 15(b)). To reduce error, we developed a map refinement procedure for edge maps. Individual links of maps can be split to a user set error threshold.

Streamwaves (Figure 15(a)). Using flow maps, we replace traditional notions of flow integration with a map lookup. This lookup enables propagation of error terms. The traditional concept of a streamline can thereby be replaced with a streamwave which propagates a fattened region forward instead of a single point, representing integration error and other measures of precision.

Comparison to other integration (Figure 15(b)). Using edge, we can compare map lookup techniques with other forms of integration. Even in simple cases, RK45 can produce inconsistent results, for example, in the spiral.

3.5 Large-Scale Parallel Morse-Smale Complex Computation

The Problem. Given the ongoing reduction in memory bandwidth compared to compute capabilities, the current strategy of data analysis as post-processing will break down. As less and less data can actually be permanently saved the analysis will become unreliable. Furthermore, any analysis will likely require the use of the same resources (e.g. super-computer) that created the



(a) A streamline using RK4 (black dashed) using stepsize $\delta t = 0.005$, Euler (black dotted) using stepsize $\delta t = 0.005$ and local exact method (LEM) (white solid), and a streamwave using edge maps with $\delta = 0.0001$ were seeded at the same point. Considering the local exact method to be the ground truth, some deviation is observed in Euler and RK4 streamlines. It is also observed that the streamwave, centered around the LEM streamline, bounds the two erroneous streamlines at all the times.

(b) Comparison between propagation using RK45 (blue) and edge maps (magenta) on a vector field defined by a counter-clockwise orbit seeded at the same point (yellow). The magenta and blue lines overlap in the beginning but a substantial deviation in RK45 streamline is observed after only one revolution around the critical point (purple). In the absence of mapping error, the mapped lines are accurate to floating-point precision.

Figure 15: Streamwaves, and comparison to other forms of numerical integration.

original data making it much more resource intensive. One potential solution to this problem is in-situ analysis performed in tandem with the simulation. This avoids repeated file I/O and since the analysis results are expected to be much smaller than the source data they can be recorded at a higher frequency. However, this framework will only be successful if the analysis can be done with minimal impact to the simulation itself.

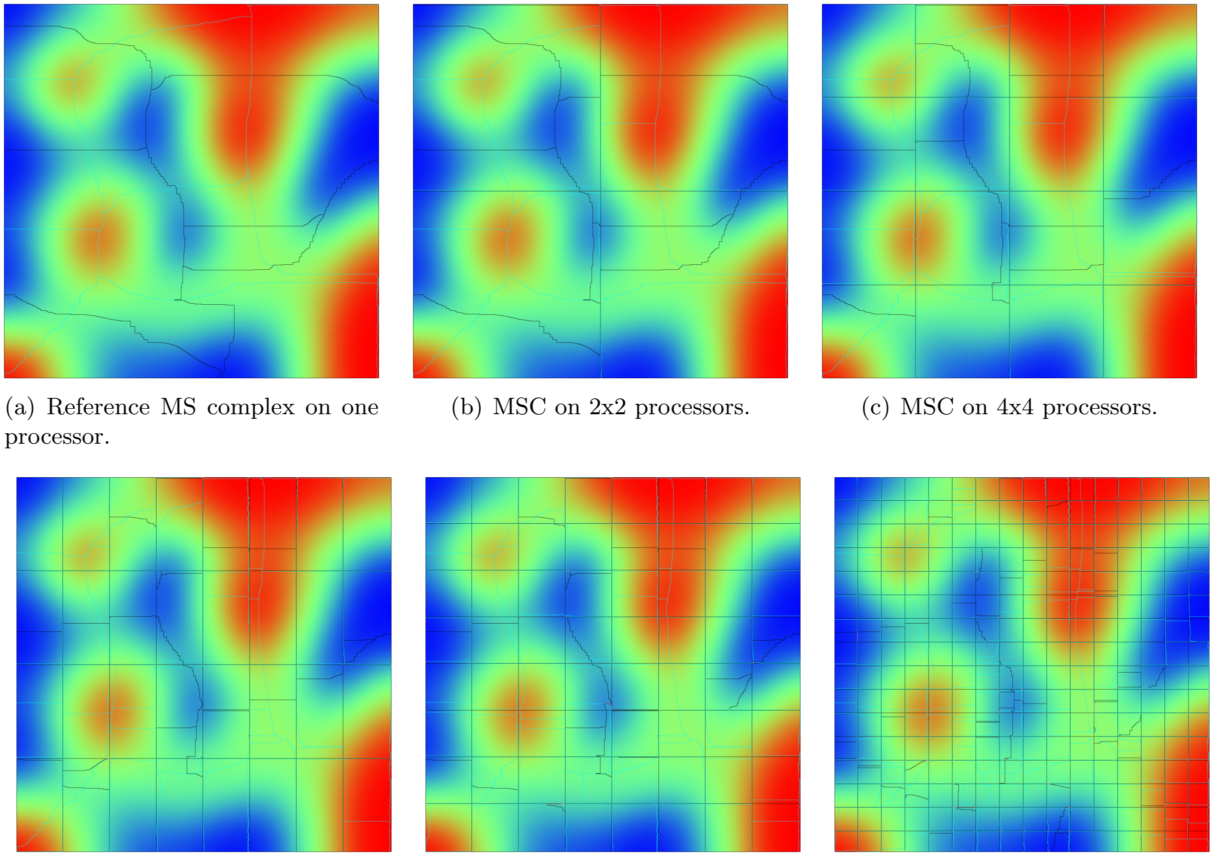
The Solution. One of the most successful and flexible data analysis frameworks is topological analysis based on Morse theory. We are developing a new massively parallel algorithm to compute, simplify, and analyze Morse-Smale complexes.

The Impact. If successful this project can provide a game-changing technology allowing detailed analysis of massive simulations. This has the potential to significantly increase our ability to extract information and insight from high resolution simulation.

Accomplishments. At this early stage, we have implemented a 2D prototype of the algorithm and have begun the study of blocking artifacts introduced by the decomposition into subdomains required for parallel computation (Figure 16).

3.6 Remote Collaboration Software Tools

Remote collaboration amongst different scientific tools enables scientists distributed across several institutions to work as a team. The main aim of our research is to define and develop an infrastructure that can enable real-time collaboration via visualization tools. Broadly speaking any viewer handles two different groups of data, the meta-data specifying viewer interaction parameters and the actual data. Two different viewers sharing and using the same meta-data is a form of bidirectional collaboration, as any viewer dependent change (meta-data) made on one of the sides will be reflected on the other. The meta-data is relatively small in size and changes dynamically. In order to tackle this real-time changing behavior, we have devised a strategy where we deal with meta-data that we represent in XML format. Having the meta-data stored in an XML tree structures gives



(a) Reference MS complex on one processor.

(b) MSC on 2x2 processors.

(c) MSC on 4x4 processors.

(d) MSC on 4x8 processors

(e) MSC on 8x8 processors

(f) MSC on 16x16 processors

Figure 16: Comparison of parallel Morse-Smale calculation at varying levels of concurrency.

us leverage both on the network front as well as the local viewer front. Using the existing XML parsing APIs and networking libraries, we are developing an real-time XML synchronization API. Using the functionalities of this API, the viewer side program can trigger appropriate callback to any remotely connected viewer corresponding to any meta-data change. Broadly speaking, the API has the following two major tasks to perform: (1) Perform local updates to the XML meta-data corresponding to any local-side viewer parameter change; (2) Corresponding to any local change initiate a callback to any remotely connected viewer by sending the updated parameter.

The Problem. DOE scientists are increasingly working in large teams distributed across the nation and the need to be able to interact more closely to share results and work together. A scientist uses a visualization tool to analyze and visualize scientific data. Usually a tool has several parameters that a scientist likes to vary, for example the viewport, dimension, color maps or any other viewer specific parameter. Our endeavor is to channel these changes to any other remote scientist working on a similar viewer and thus enable remote communication between them. The major bottleneck lies in handling/synchronizing dynamically changing data.

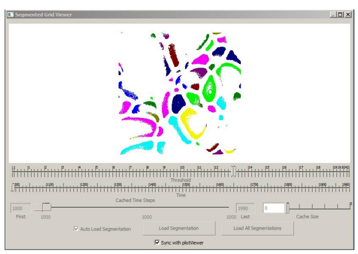
The Solution. We are developing an API that can achieve the desired effects overcoming the synchronization problem. The API will have the following two features that will help in managing dynamically changing large meta-data files: (1) In-memory Meta-Data – using this functionality of the API, one can load the XML files in memory and save on any I/O operations, which will address issues related to large-sized meta-data files; (2) In File Meta-Data – instead of loading the meta-data in memory, this method will actually write meta-data into XML files and perform the

necessary actions on them.

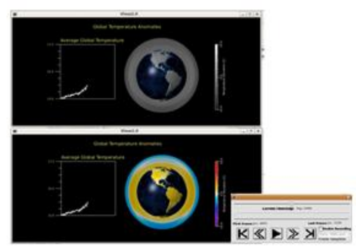
The Impact. We are still in the development phase of this project, though progress made over the past few months gives us enough confidence that the collaborative tools will greatly aid people with similar research interests. Researchers using the tools will be able to remotely interact with one other and collaborate their work. Our two target applications are tools for combustion scientists (combustion end-station, Figure 3.6(a)) and climate modeling via CDAT (Figure 3.6(b)). We are also using a third application as a pilot testbed for remote manual segmentation of microscopy data (Figure 3.6(c)).

Accomplishments. Our goal is to achieve a distributive collaborative environment, where we can have a server as the governing/controlling entity and we can have subsequently many clients connected to the server that can remotely collaborate with each other.

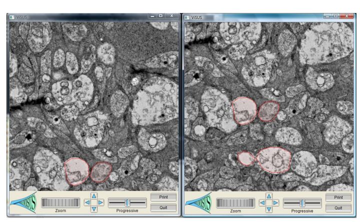
- Design of a prototype API that can synchronize XML files remotely, enabling the API to handle both in memory as well as a physical file for XML meta data. The API initiates client side communication with the remote server, which also interacts with other clients.
- Initial implementation was that of a push server, that required constant querying on the part of the clients, leading to obvious waste in time and computation resources. We have successfully managed to change the design of the server from being a push server to a pull/daemon server, leading to a considerable improvement in performance.



(a) Segmentation Viewer: For visualizing Combustion data. Our target is to add collaborative features to the viewer.



(b) Tool for analyzing climate data, we have partially incorporated collaborative features to the viewer.



(c) Visus 2D IDX viewer – We have attained collaboration for the earlier version of IDX viewer, out target is to attain collaboration for the current versions.

4 Common Infrastructure Projects

4.1 VisIt Infrastructure: Data Subset Selections

The Problem. Effective techniques for large data exploration and analysis consist of a mixture of solutions. One dimension of this problem is to effectively select subsets of large datasets for subsequent exploration and analysis. Software tools that process very large data often have trouble sifting data efficiently to isolate features, or data subsets, of interest. This can be especially true of time-evolving particle-based data sets where particles come and go over time.

The Solution. Our solution to helping users isolate features of interest, specifically collections of particles in time evolving particle-based data, is to introduce selections to VisIt. VisIt's various methods of dataset reduction (threshold, slicing, etc.) are applied to a plot and the plot's output may constitute a selection, a reduced set of cells selected from the input dataset. The selection can be applied to other plots, making it easy to look at different views of the data. The selection is linked to plots that use it so when the plot that created the selection updates, the dependent plots also update. Selections are also used in conjunction with selection-aware database readers such as VisIt's H5Part reader in order to limit the amount of data when it is read from the file. This optimization saves memory and time spent doing I/O.

We have exposed the work on selections via VisIt's user interface, giving even novice VisIt users the ability to reduce the amount of data being processed and to work with multiple plots linked via selections.

Now that the first round of selection work has been added to VisIt's user interface, we plan to build on that and add support for *cumulative queries*. A cumulative query is a selection that is in some ways similar to a threshold operation in that it permits you to restrict the dataset based using a compound boolean range query, i.e., several ranges over one or more variables. A cumulative query differs in its ability to combine the results of the threshold selection for a range of time steps in a time-varying dataset. This enables the user to create selections that fit the range criteria for ANY or ALL selected time steps, making it easier to track time-varying phenomena using selections.

The Impact. VACET has added many enabling technologies for fusion scientists. With minimal help, VisIt user Charlson Kim from the University of Washington was able to produce a fairly sophisticated visualization using, among other features, streamlines and selections (Figure 17).

> I have been waiting for a tool like this (VisIt) all my life.
>
> –Charlson Kim, University of Washington

Accomplishments. The initial selection implementation exposed a minimal Python-only interface that allowed users to create and apply selections in VisIt's compute engine. We have taken this much farther and exposed selections in the VisIt GUI and viewer, making them an accessible part of VisIt. VisIt's GUI and viewer now have a notion of a selection list that contains properties about all of the selections in existence. This allows VisIt's GUI to provide controls for managing selections as it is able to observe state from the viewer. The GUI was also enhanced so that a plot's entry in the plot list indicates whether it creates or uses selections. VisIt's viewer was enhanced so plots are selection-aware, meaning that when a plot creates a selection and gets updated, other plots that use the selection are also updated. Additional measures were taken to ensure that selections can be recreated appropriately when plots are created as a result of restoring a VisIt session. This work was deployed in VisIt's 2.1 release.

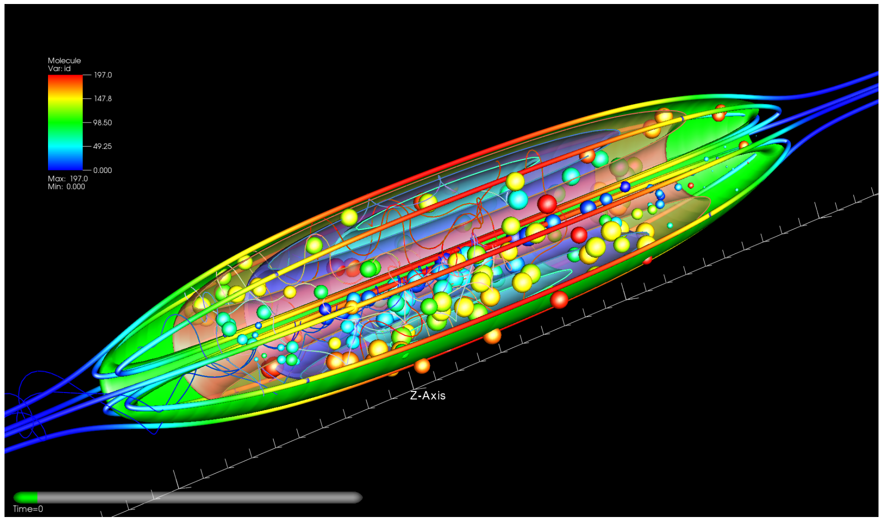


Figure 17: This image shows a typical Field Reverse Configuration (FRC) with particle trajectories. Charlson Kim works on hybrid PIC-MHD simulations that combine effects of fluid simulations and PIC simulations. The image shows tracer PIC trajectories in an static equilibrium plasma configuration called the FRC. In the image, the fluid part of the plasma is visualized with the streamlines and contours. The streamlines are magnetic fields colored by their magnitude (two sets are displayed, a set representing the last closed field lines (forming closed loops), and a set representing the first open field line (leaving on either side of the images). The contours show pressure, increasing in pressure (equivalently, temperature) moving inwards. The particles are the spheres and squiggly traces. The color is just the ID, size of the sphere is proportional to the perpendicular component (perpendicular to magnetic field) of the kinetic energy. The squigglies show a subset of the trajectories. The trajectories directly relate to confinement, the goal of fusion. The image show a highly idealized scenario. Image courtesy of Charlson Kim, PSI Center, Univ. of Washington, CEMM SciDAC.

Figure 18 illustrates this capability. A user selects a subset of a time-varying particle-based data set via a multivariate range query specified via a parallel coordinates interface. VisIt then selects the data subset and visualizes it in the other window.

4.2 VisIt Multiresolution Support

The Problem. The results of large scale simulations are generating ever-larger data sets, outpacing traditional desktop hardware advances and even medium scale supercomputing resources. This adds a logistical challenge – acquiring dedicated large scale computing resources – to the tasks of visualization and analysis, and therefore novel scientific insight. Even at the largest scale, complicated visualizations can take significant computational resources and time. Yet interactive feedback is critical to encourage and adopt an exploratory process that will lead to novel scientific insight, as opposed to reinforcing established knowledge.

The Solution. In many cases, viewing a lower resolution version of the data is sufficient for observing major trends and characteristics of a calculation. This is the central thesis behind adaptive mesh refinement (AMR), for example: use coarser grid resolutions where the impact of such a discretization will be negligible. This approach can save significant computational resources that can be applied elsewhere, or, in the case of visualization and analysis, simply accelerates the feedback from interaction to result.

We provide a mechanism for a data format to inform VisIt of its multiresolution structure. The user may then instruct VisIt to process lower, or higher resolution data as desired. Processing

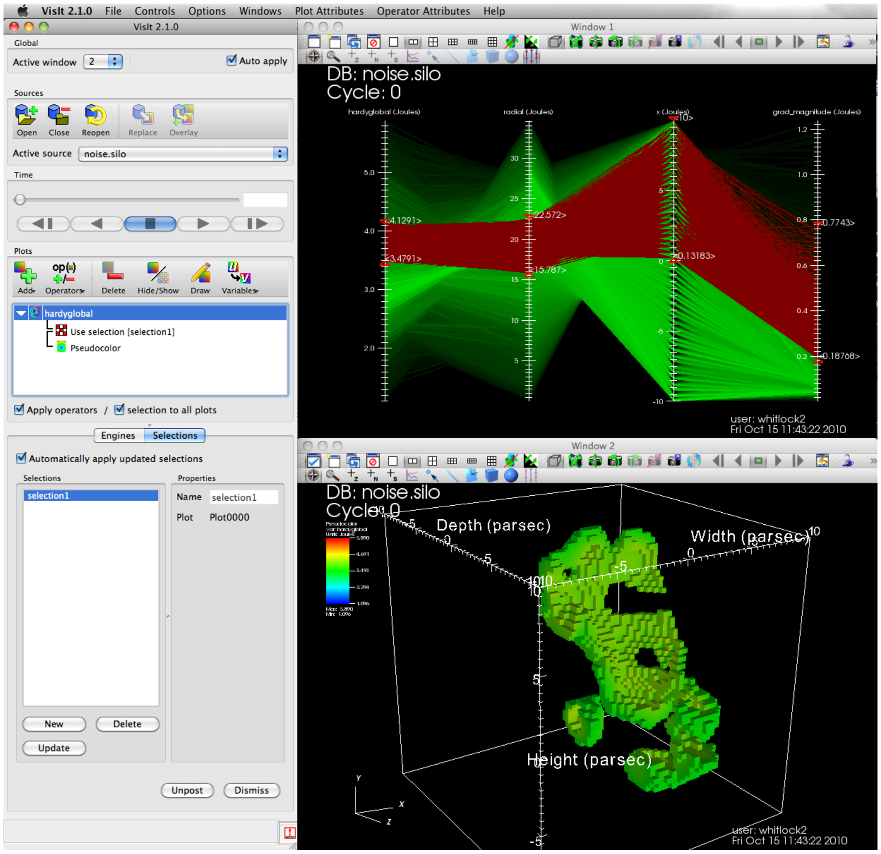
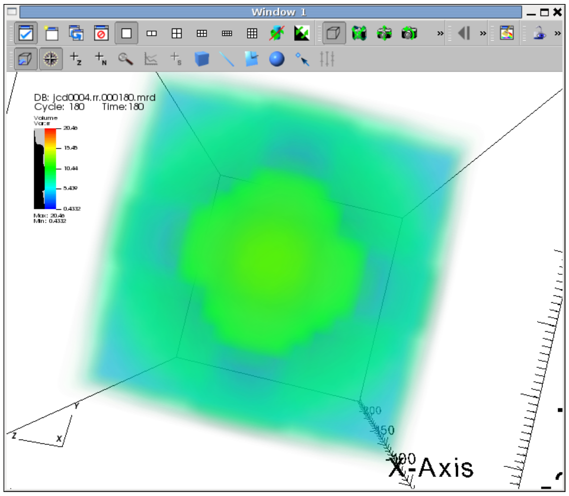


Figure 18: A user selects a subset by specifying a multivariate range query via a parallel coordinates plot (top). Then, VisIt extracts this subset, making it available to any active plots for visualization (bottom).

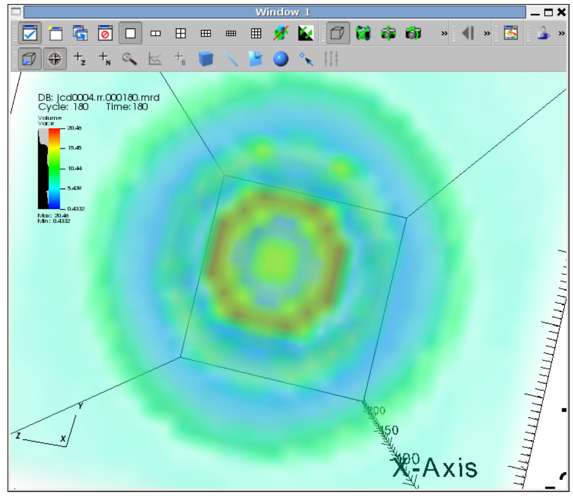
lower-resolution versions saves computational time and resources. The system allows a user to easily load up lower resolutions of their data on a standard workstation or laptop, priming their visualization tasks for when supercomputing resources are available. It helps them to quickly obtain fast feedback on a visualization to explore the best parameters to use for larger, more expensive runs using full resolution versions of the data.

The Impact. Multiresolution support in a visualization and analysis tool allows and encourages exploratory visualization. As data sizes increases, generating informative visualizations from the data becomes a correspondingly more difficult problem. The effect is to "bake" existing visualization tasks into a workflow and re-execute it as new data become available. While this process has its place, novel insights concerning phenomena that emerge at scale and were not predicted a priori will be missed. For this reason an exploratory analysis process must be available, through which unexpected features can be discovered and understood.

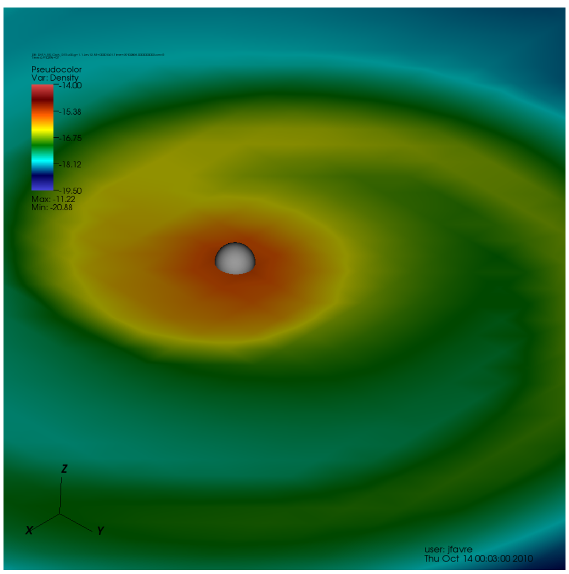
> [...] the advantage is between "sluggish rendering" and "interactive rendering." From the developer's point of view, the amount of work was "10 minutes, 10 lines of code."
>
> –Jean Favre, Swiss National Supercomputing Center.



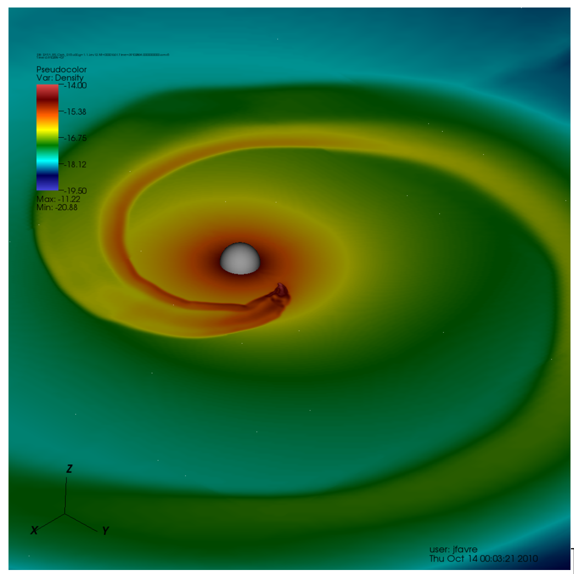
(a) Using the 'multires' operator to select a low resolution version of a data set.



(b) The same data set rendered at native resolution.



(c) Using the 'multires' operator to select a low resolution version of a data set generated at the Swiss National Supercomputing Centre. Image courtesy J. Favre, SCSC.



(d) The same data set rendered at native resolution. Image courtesy J. Favre, SCSC.

Figure 19: Example VisIt output comparing visual output created using lower- and higher-resolution data.

Accomplishments. We have implemented a multiresolution solution inside VisIt. Supported formats expose a resolution widget that allows interactive switching between available resolutions. VisIt utilizes the information given in the UI to avoid selecting resolutions that are not desired; such data are never read from disk, significantly reducing overhead. Both AMR and multiresolution data are supported. Example visual output from this new functionality is shown in Figure 19.

Support must be specifically written into a VisIt database loader for this infrastructure to work. Currently the Chombo and STAR file formats include such support. We have also created exhaustive documentation on how to enable support for the required contract, and ensured only minimal changes to the database loader are needed to enable support for multi-resolution loading,

analysis and exploration.

4.3 VisIt Releases and Release Engineering

The Problem. We are improving the software development process for VisIt, which is one of the tools used to deploy VACET technology, to better serve VACET customers. Specifically:

- We want to be able to deploy new functionality to VACET customers more quickly.
- We want to let VACET customers know what new functionality is being developed and when it will be released to them.
- We want VACET customers to be able to submit their own bug reports and feature requests and give them timely feedback as to progress being made on the bug report/feature request and when it will be available to them.

The Solution. We have replaced the existing bug and feature tracking system that was only accessible to developers with accounts on LLNL computer accounts with one that is now available to anyone in the world. We have once again moved to a more frequent and predictable release cycle for VisIt. We are tracking more of our software development in a new bug and feature tracking system.

The Impact. The new bug and feature tracking system allows customers to enter their own bug reports and feature requests, track the progress of the bug report/feature request, and find out when it will be available to them. By moving to a more frequent release schedule customers will receive new functionality quicker. A side benefit will be higher quality code, since the less frequent the release schedule the higher the likelihood that developers will try add new features at the last moment before a release, which results in less well tested code with a higher likelihood of bugs. Tracking all of the software development through the bug and defect tracking system allows better tracking of the progress of the releases and increases the likelihood of the software being released on schedule.

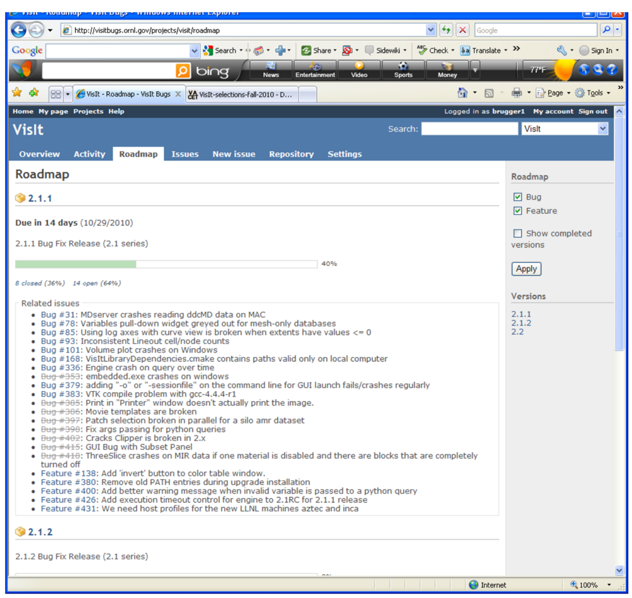


Figure 20: Screen shot of the new VisIt bug tracking system.

Accomplishments.

- We have deployed the Redmine bug and feature tracking system. It has been deployed at Oak Ridge National Laboratory and is accessible to anyone in the world once they register on the system (Figure 20).
- We have moved to a more frequent release schedule with major releases scheduled 3–4 months apart and 1–2 patch releases between each major release.
- We are tracking about 50% of the development happening in VisIt through the new Redmine bug and feature tracking system. It has a roadmap feature which has dates of upcoming releases as well as a list of all the bugs and features targeted for a particular release. The list of bugs and features also has the status of each of the bugs and features, indicating if it is completed or not.

4.4 High Performance Data Analysis

Problem. Statistical analyses (clustering, regression, extremes modeling, sampling, analysis of variance, etc.) are not available for large data. As a result, statistical analyses are used in limited ways in large data visualization.

Solution. Our approach is to enable the R software environment for statistical computing and graphics[5] to be used for data-parallel analysis on DOE's large computational platforms. We also aim to make data-parallel R methodology available to VisIt visualization. We first focus on analysis of extremes and clustering, which are methods immediately relevant to out stakeholders.

Impact. Enable scientists to see through the "fog of variability" in large data sets. Statistical methods are descriptions of variability and its underlying governing principles. Facilitate visualization at petascale and beyond by statistical selection of representative features. This is primarily accomplished through clustering of features into a smaller number of classes. A broader impact is to bring applied mathematics (represented by statistics) as players to the large data analysis table and develop more large data statistical analysis algorithms.

Accomplishments this Period.

Much work in this period centered around the application of the k-means clustering code to climate data. The code uses uncertainty quantification in two important ways. First, estimation of the k means only requires a sample of the data for sufficient precision, thus providing an order of magnitude run time speedup. Second, classification uncertainty measured by the level of agreement among the random starts provides the selection criterion for choosing k to be the best number of clusters supported by the data. The specific agreement computation chosen accounts for expected agreement in a way that is appropriate for highly unequal cluster sizes and massive data sets.

We used a 3 GB climate data set that presents a moderate clustering problem with 120 million objects in a 5 dimensional space consisting of large scale precipitation, convective precipitation, surface pressure, surface temperature, and temperature at reference height. The agreement matrix among 8 random starts indicates that five clusters are best supported by the data as shown in Figure 21.

A set of 8 random starts for a given k on 256 processors takes about 2 minutes to complete. To visualize the clustering results, we output a NetCDF file of the same mesh as the input variables and use VisIt to play back the cluster evolution along with the variables. A snapshot of the movies is shown in Figure 22.

[5] See `http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html`.

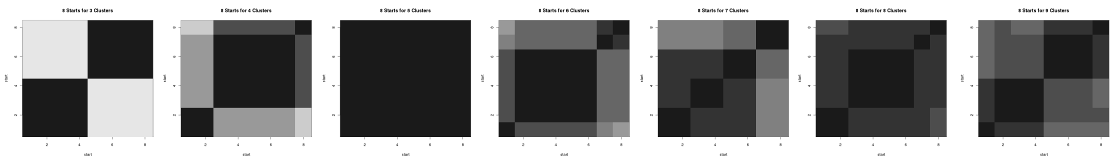


Figure 21: Agreement between 8 random starts for 3 through 9 k-means clusters (dark is high agreement). Data best supports five clusters.

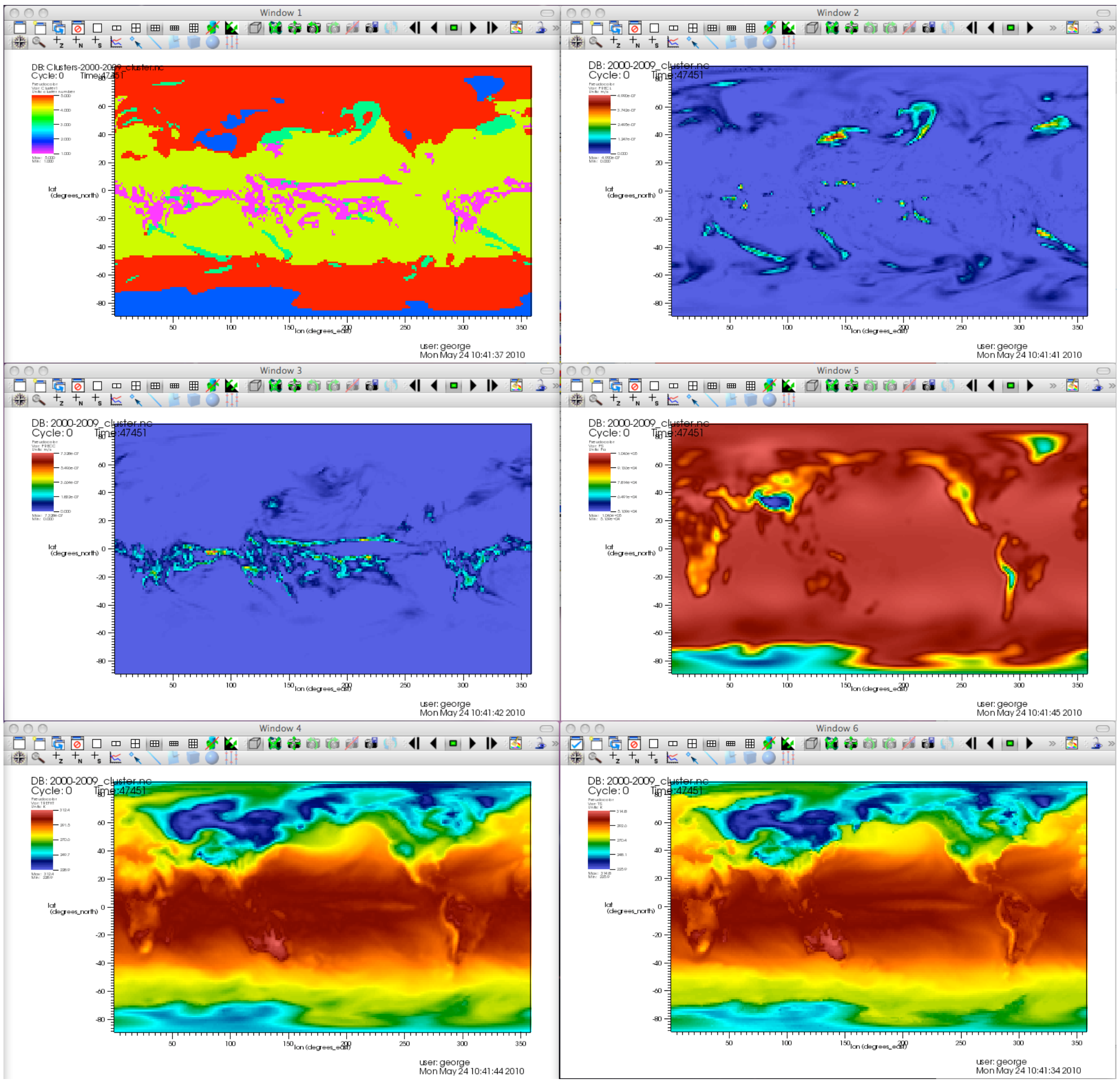


Figure 22: Clockwise from top right we have, *large scale precipitation*, *surface pressure*, *temperature at reference height*, *surface temperature*, *convective precipitation*, and the *five computed clusters*. The clusters have picked up the large scale precipitation (light green) and convective precipitation (magenta). Here, clustering is being used as a tool for automated feature identification and tracking. Data without spatial or time information is clustered and then displayed in space and time, revealing and tracking the identified features.

4.5 ViSUS Infrastructure

The Problem. The ability to visualize and analyze simulations in-situ can enable stakeholder scientists provide feedback to help guide the simulation, or take necessary action if unexpected

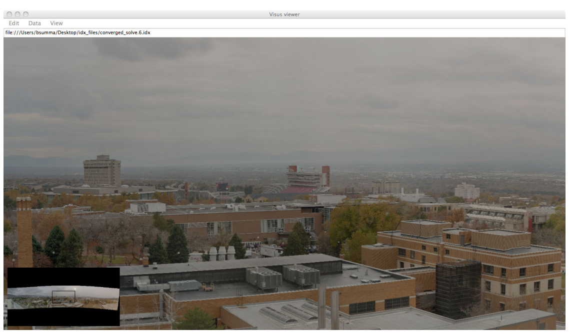
results emerge, saving precious time and resources for large simulation runs. Moreover, such in-situ visualizations and analysis capabilities must not add excess overhead to the resources required for the simulation. The true power of these new capabilities is only realized when scientists are able to visualize and analyze data remotely, in collaboration with their colleagues and on a variety of platforms, whether on a Powerwall or an iPhone. Our specific target is to provide combustion scientists with a remote, in-situ, collaborative infrastructure.

The Solution. With its new stable IO infrastructure, we've moved ViSUS forward as a solution to this problem. The work since last spring has been on two fronts:

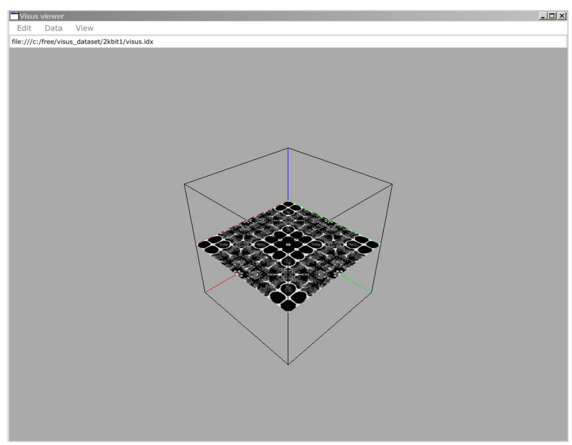
- The ViSUS viewer application had, over the years, become highly customized making extending and porting the viewer non-trivial. We've redesigned the viewer from scratch to alleviate these problems. We've also added all the needed hooks to enable future integration with the remote collaboration system also being developed by the group.
- In visualization researchers at Argonne (Carns, Latham, Peterka, Ross, Vishwanath, Papka), there has been significant progress in the development of a parallel ViSUS I/O infrastructure. The intent in the long term is to have these routines can be used in-situ to write simulation data directly in the ViSUS IDX format enabling remote viewing by application scientists.

The Impact. We are still much in the development phase of this project, though progress made over the past few months gives us confidence that the ViSUS viewer and infrastructure will reach a distributable state in the near future.

Accomplishments.



(a) Sample output from the new 2D ViSUS Viewer.



(b) Sample output from the new 3D ViSUS Viewer.

Figure 23:

ViSUS and Collaborative Data Viewer

Main additions to ViSUS I/O:

- New SWIG wrapper for python and C#.
- Standard ViSUS sub-sampled hierarchy replaced with user defined filters. Three standard filters: Haar wavelets, subsampling, and Min/max are provided as defaults.
- Added FreeImage library to support multiple image formats for both input and output.
- Preliminary work on lossy data compression for IDX.

The ViSUS viewer has been completely redesigned to allow for a highly portable and extensible application (Figure 23). New features include:

- Single viewer supports both 2D data, 3D data and either over time.
- OpenGL and OpenGL ES support for portable rendering.
- Viewer designed with multiple platforms in mind. We can now build and distribute the viewer on multiple platforms (including iPhone and iPad) from a single source. (this also gives us an added benifit of supporting gestures in all versions of the viewer, which will be helpful when porting to touchscreen systems).
- Viewer can also be used as a data server or a remote client.
- Lossy or lossless compression can be enabled for remote data visualization.
- Progressive volume rendering/slices/isosurfaces for 3D data.
- Preliminary work on a client for remote data visualization via a web browser.
- PiP (Picture-in-picture) feature added to give a coarse/global context when exploring large datasets.

Parallel ViSUS

Our preliminary results demonstrate 60% of the peak I/O throughput when reorganizing and writing the data from 512 processes on an IBM BG/P system.

Serial writer. The first step was to create parallel application that would use ViSUS I/O to write directly into IDX format. Each process writes independently to an IDX data set and MPI barriers/tokens were used to maintain order amongst processes. The processes cannot write concurrently due to conflicts in updating metadata and block layouts. The efficiency improves as the aggregate volume to write is increased to 8 GB, but its maximum performance is only 9.5 MB/s. We obtained a peak performance of 218MB/s for 64 processes writing a total of 8GB. Thus, we only achieve about 4% of the maximum throughput with the serial writer. This is to be expected as ViSUS is serial in nature and the various processes take turns to write the data out.

Parallel writer. Based on our experience with the serialized ViSUS writer, we then developed a prototype API for performing concurrent I/O to an IDX data set. This API, called Parallel IDX (PIDX), includes functions patterned after ViSUS for creating, opening, reading, and writing IDX data sets. Each of the PIDX functions is a collective operation that accepts an MPI communicator as an argument.

Optimization. We implemented a file handle caching mechanism in our prototype. When any process opens an underlying binary file, it holds the MPI file descriptor open for future use and does not close it until all I/O is complete. We notice a significant improvement of up to 7-fold for data volumes of 128MB. However, we also had only a marginal improvement with higher data volumes. This was due to the significant amount of the I/O time that was spent in the computation to generate the Hierarchical-Z ordering. We performed a detailed analysis of the HZ computation using the IBM BG/P universal performance counters and identified bottlenecks associated with redundant computations as well as inadequate use of the floating point double numbers. By overcoming these, we were able to achieve up to 75% improvement in I/O throughput over the file handle improvements and up to a 10-fold improvement over the default implementation.

4.6 Collaboration Summary

This list reflects collaborators from this reporting period.

SciDAC Science Applications

- Stephane Ethier, Ku, Charlson Kim (fusion, particles, query stuff).
- Scott Kruger (Tech X) and Josh Breslau (PPPL) of the CEMM SciDAC (Jardin – PI), collaborators on the Poincaré analysis software deployed in VisIt (Section 2.7).

- Bill Nevins (LLNL) and Jeff Candy (GA) of the Micro Turbulence SciDAC (Bill Nevins – PI), collaborators on the Poincaré analysis software deployed in VisIt (Section 2.7).
- Raul Sanchez and Don Bachelor (ORNL) of the CSWIM SciDAC (Don Bachelor – PI), collaborators on the Poincaré analysis software deployed in VisIt (Section 2.7).
- SciDAC CAC, Stan Woosley, Jason Nordhaus.

SciDAC Centers

- SciDAC Scientific Data Management Center. We are using FastBit as the basis for index/query technolgy that is integral to many of our projects, like the fusion particle analysis work (Section 2.6) that is built upon persistent, cumulative queries (Section 4.1).

Other Computer Science Researchers

- An item.

4.7 Resources

Blurb about compute time at NERSC and ORNL.

5 Publications

5.1 Peer-reviewed Journal Articles

1. John C. Anderson, Christoph Garth, Mark A. Duchaineau, and Ken Joy. Smooth, Volume-Accurate Material Interface Reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 16(5), 2010.

2. P.-T. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. B. Bell. Interactive Exploration and Analysis of Large Scale Simulations Using Topology-based Data Segmentation. *IEEE Trans. on Visualization and Computer Graphics*, page to appear, 2010.

3. Christoph Garth and Kenneth I. Joy. Fast, memory-efficient cell location in unstructured grids for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 2010. To appear.

4. H. Childs, D. Pugmire, S. Ahern, B. Whitlock, M. Howison, Prabhat, G. Weber, and E. W. Bethel. Extreme Scaling of Production Visualization Software on Diverse Architectures. *Computer Graphics And Applications, Special Issue on Ultrascale Visualization*, 30(3):22–31, May/June 2010.

5. L. J. Gosink, C. Garth, J. C. Anderson, E. W. Bethel, and K. I. Joy. An Application of Multivariate Statistical Analysis for Query-Driven Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 99, 2010. RapidPosts, LBNL-3536E.

6. Mathias Hummel, Christoph Garth, Bernd Hamann, Hans Hagen, and Kenneth I. Joy. IRIS: Illustrative Rendering of Integral Surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 2010. To appear.

7. M. Isenberg, P. Lindstrom, and H. Childs. Parallel and Streaming Generation of Ghost Data for Structured Grids. *Computer Graphics And Applications, Special Issue on Ultrascale Visualization*, 30(3):50–62, May/June 2010.

8. Allen R. Sanderson, Guoning Chen, Xavier Tricoche, David Pugmire, Scott Kruger, , and Joshua Breslau. Analysis of Recurrent Patterns in Toroidal Magnetic Fields. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 2010.

9. Brian Summa, Giorgio Scorzelli, Ming Jiang, Peer-Timo Bremer, and Valerio Pascucci. Interactive Editing of Massive Imagery Made Simple: Turning Atlanta into Atlantis. *ACM Transactions on Graphics - in press*, page 12, 2010.

5.2 Peer-reviewed Conference Proceedings

1. Thomas Fogal and Jens Krüger. Tuvok, an Architecture for Large Scale Volume Rendering. In *Proceedings of the 15th International Workshop on Vision, Modeling, and Visualization*, November 2010. to appear.

2. Thomas Fogal, Hank Childs, Siddharth Shankar, J. Krüger, R. Daniel Bergeron, and Philip Hatcher. Large Data Visualization on Distributed Memory Multi-GPU Clusters. In *Proceedings of High Performance Graphics 2010*, pages 57–66, June 2010.

3. F. Fuentes, H. Kettani, G. Ostrouchov, M. Stoitsov, and H. A. Nam. Exploration of high-dimensional nuclei data. In *International Conference on Communication Software and Networks*, pages 521–524. IEEE, 2010.

4. Mark Howison, E. Wes Bethel, and Hank Childs. MPI-hybrid Parallelism for Volume Rendering on Large, Multi-core Systems. In *Eurographics Symposium on Parallel Graphics and Visualization (EGPGV)*, Norrköping, Sweden, May 2010. LBNL-3297E, Best Paper Award.

5. R. Kettimuthu, A. Sim, D. Gunter, B. Allcock, P.-T. Bremer, J. Bresnahan, A. Cherry, L. Childers, E. Dart, I. Foster, K. Harms, J. Hick, J. Lee, M. Link, J. Long, K. Miller, V. Natarajan, V. Pascucci, K. Raffenetti, D. Ressman, D. Williams, L. Wilson, and L. Winkler. Lessons learned from moving earth system grid data sets over a 20 gbps wide-area network. In *Proc. ACM International Symposium on High Performance Distributed Computing, HPDC, 2010*, pages 316–319, 2010.

6. J. S. Meredith and H. Childs. Visualization and Analysis-Oriented reconstruction of material interfaces. *Computer Graphics Forum (Proceedings of Eurovis 2010)*, 29(3):1241–1250, 2010.

7. title = Two-stage Framework for a Topology-Based Projection and Visualization of Classified Document Collections P. Oesterling and G. Scheuermann and S. Teresniak and G. Heyer and S. Koch and T. Ertl and G. H. Weber. In *Proceddings IEEE Symposium on Visual Analytics Science and Technology (IEEE VAST)*, October 2010.

8. Oliver Rübel, Sean Ahern, E. Wes Bethel, Mark. D Biggin, Hank Childs, Estelle Cormier-Michel, Angela DePace, Michael B. Eisen, Charless C. Fowlkes, Cameron G. R. Geddes, Hans Hagen, Bernd Hamann, Min-Yu Huang, Soile V. E. Keränen, David W. Knowles, Cris L. Luengo Hendriks, Jitendra Malik, Jeremy Meredith, Peter Messmer, Prabhat, Daniela Ushizima, Gunther H. Weber, and Kesheng Wu. Coupling Visualization and Data Analysis for Knowledge Discovery from Multi-dimensional. In *Procedia Computer Science, Proceedings of International Conference on Computational Science, ICCS 2010*, June 2010. LBNL-3669E.

9. K. Potter, J.M. Kniss, R. Riesenfeld, and C.R. Johnson. Visualizing Summary Statistics and Uncertainty. *Computer Graphics Forum (Proceedings of Eurovis 2010)*, 29(3):823–831, 2010.

10. E. Deines, G. H. Weber, C. Garth, B. Van Straalen, S. Borovikov, D. F. Martin, and K. Joy. On the Computation of Integral Curves in Adaptive Mesh Refinement Vector Fields. In *Proceedings of Dagstuhl Seminar on Scientific Visualization 2009*, 2010.

11. Jian Cui, Paul Rosen, Voicu Popescu, and Christoph Hoffmann. A curved ray camera for handling occlusions through continuous multiperspective visualization. In *IEEE Visualization 2010*, 2010.

12. S. Kumar, V. Pascucci, V. Vishwanath, P. Carns, R. Latham, T. Peterka, M. Papka, and R. Ross. Towards parallel access of multi-dimensional, multi-resolution scientific data. In *5th Petascale Data Storage Workshop (at Supercomputing 2010)*, 2010.

5.3 Posters

1. Mark Howison, David Camp, Dave Pugmire, Christoph Garth, Hank Childs, Ken Joy, and E. Wes Bethel. Hybrid-parallelism Improves Visualization Performance on Large, Multi-core Systems. In *Scientific Discovery through Advanced Computing 2010*, Chattanooga, TN, July 2010.

2. Christoph Garth, Hari Krishnan, Ken Joy, and Hank Childs. Advanced Vector Field Analysis Using Integration-based Visualization. In *Scientific Discovery through Advanced Computing 2010*, Chattanooga, TN, July 2010.

5.4 Book Chapters

1. Christopher Co, Mark A. Duchaineau, and Kenneth I. Joy. Streaming aerial video textures,. In H. Hagen, editor, *Scientific Visualization: Advanced Concepts*. Schloss Dagstuhl, Leibnitz Center for Informatics, 2007.

6 Outreach, Service, and Awards

6.1 Outreach

Invited Talks

- T. Fogal, "Parallel Volume Rendering on GPU Clusters." FACETS Meeting, Boulder, CO. Aug 18-20, 2010.
- O. Ruebel, "Analysis of Particle Beams in Laser Wakefield Particle Acceleration Data," International Research Training Group 1131, University of Kaiserslautern, Germany, June 7, 2010.
- G. Ostrouchov, "Data-parallel statistical computing: a model based clustering example," Joint Research Conference on Statistics in Industry and Technology, Gaithersburg, Maryland, May 25-27, 2010.
- G. Ostrouchov, "Parallel statistical computing: Are we embracing the scalable concurrency revolution?" Joint Statistical Meetings, Vancouver, Canada, August 1-6, 2010.
- E. Wes Bethel, "MPI-hybrid Parallelism for Volume Rendering on Large, Multi-core Systems." Astronum 2010, 17 June 2010, San Diego CA.
- H. Childs, "Why Large Scale Visualization & Analysis Will Change the Rules," Florida State Scientific Computing Colloquium, Tallahassee, FL, October 13, 2010.

- H. Childs, “Overview of VisIt,” MIT Fusion science seminar. Cambridge, MA, September, 2010.
- H. Childs, “Parallel Particle Advection,” CScADS 2010, Snowbird, UT, July, 2010.
- H. Childs, “Overview of VisIt,” CScADS 2010, Snowbird, UT, July, 2010.
- H. Childs (on behalf of M. Howison and E.W. Bethel), “MPI-hybrid Parallelism for Volume Rendering on Large, Multi-core Systems,” Norrköping, Sweden, May 2010.
- H. Childs, “Using VisIt to Visualize Turbulent Data,” NORDITA and Linne FLOW Centre Workshop on Turbulent Boundary Layers, Stockholm, Sweden, April, 2010.
- D. Pugmire, “Parallel Visualization Scalability,” CScADS 2010, Snowbird, UT, July, 2010.
- G. H. Weber, “VACET: Deploying Technology for Visualizing and Analyzing Astrophysics Simulations,” SciDAC Computational Astrophysics Consortium Meeting, Menlo Part, CA, May 2010.
- G. H. Weber, “Topology-based Feature Definition and Analysis,” Astronum 2010 – Fifth International Conferece on Numerical Modeling of Space Plasma Flows, San Diego, CA, June 2010.
- C. Johnson, “Visual Computing: Making Sense of a Complex World,” Bucknell University, Lewisburg, October 2010 (Emerging Minds Series).
- C. Johnson, “Large-Scale Visualization,” IEEE Cluster 2010, Crete, Greece, September 2010 (Keynote Presentation).
- C. Johnson, “Extreme-Scale Computational Biomedicine,” ICIS Workshop: Future of the Field, August 2010.
- C. Johnson, “Image-Based Biomedical Modeling, Simulation, and Visualization,” Wright State University, Dayton, May 2010 (2010 Mazumdar Lecture in Applied Mathematics).
- C. Johnson, “Image-Based Biomedical Modeling, Simulation, and Visualization,” Mathematical Biosciences Institute, Ohio State University, May 2010.

Tutorials

- H. Childs, “VisIt Overview,” US Army Research Lab, Aberdeen, MD, September 2010.
- H. Childs and D. Pugmire, “VisIt Overview,” SciDAC 2010, Chattanooga, TN, July 2010.
- K. Joy, “Material Interface Reconstruction,” Los Alamos National Laboratory, Los Alamos NM, August 2010.
- K. Joy, “Material Interface Reconstruction,” Technical University of Kaiserslautern, Kaiserslautern Germany, August 2010.
- K. Joy, “Material Interface Reconstruction,” Leipzig University, Leipzig Germany, August 2010.

Symposia and Workshops

- H. Childs, Participant in “NSF Workshop on Software Development Environment for Science & Engineering Applications,” September, 2010, Washington DC.

6.2 Service

Program Committee

- G. Ostrouchov, Section Program Chair, Joint Statistical Meetings (Section on Physical and Engineering Sciences), August 1-6, 2010, Vancouver, Canada.

- C. Johnson, Selection Committee, SIAM/ACM Prize in Computational Science, 2010.
- C. Johnson, Program Committee, Parallel Processing for Imaging Applications, 2010.
- C. Johnson, Program Committee, Workshop on Novel Computing for Life Sciences, 2010.
- K. Joy, Program Committee, IEEE Visualization 2010.
- K. Joy, Program Committee, TopoInVis 2010.
- K. Joy, Program Committee, Interactive 3D Computer Graphics and Games.

Reviewer

- The Visual Computer Journal.
- IEEE Visualization 2010 Technical Program.
- IEEE InfoVis 2010 Technical Program.
- Annual Workshop of the International Research Training Group 1131, 2010, Technical Program.
- EuroVis 2010 Technical Program.
- DOE SBIR program, 2010.
- NSF Software Infrastructure for Sustained Innovation (SI^2).
- IEEE Transactions on Visualization and Graphics.
- IEEE Visual Analytics for Science and Technology (VAST) Technical Program.
- ACM Transactions on Graphics.
- Journal of Engineering with Computers
- Pacific Visualization 2011 Technical Program.
- Astronum 2010 Technical Program.

Advisory Panels

- C. Johnson, NSF Office of Cyberinfrastructure Task Force on Software Infrastructure.
- C. Johnson, NSF Office of Cyberinfrastructure Task Force on Grand Challenge Communities.
- C. Johnson, CRA Computing Community Consortium (CCC).
- C. Johnson, Computing Research Association Education Committee.
- C. Johnson, Fundamental and Computational Science Directorate Review Committee, Pacific Northwest National Laboratory.
- C. Johnson, European Union Science Foundation, Virtual Physiological Human Network of Excellence, International Advisory Board.
- C. Johnson, NIH National Center for Biomedical Computation, Stanford University, Scientific Advisory Board.
- C. Johnson, Finnish Centre of Excellence in Inverse Problems, International Scientific Advisory Board.
- C. Johnson, Bavarian Graduate School of Computational Engineering, International Advisory Board.
- C. Johnson, NIH National Alliance for Medical Image Computing, Advisory Board.
- K. Joy, External Advisory Board, NIH Center for Integrative Biomedical Computing, University of Utah.

Editorial Boards

- C. Johnson, Journal of Uncertainty, Editorial Board.
- C. Johnson, Journal of Computational Science, Editorial Board.

- C. Johnson, DOE Office of Advanced Scientific Computing Research Communications Project Editorial Board.
- C. Johnson, Journal of Computing and Visualization in Science, Editorial Board.
- K. Joy, Graphical Models, Editorial Board.

6.3 Awards

1. George Ostrouchov, Named 2010 Fellow of the American Statistical Association, "For excellent and sustained research and collaboration involving the statistical analysis of massive data sets," which in part stems from participation in the VACET project collaboration.

2. Chris Johnson, Received the 2010 Rosenblatt Prize for Excellence, the University of Utah's most prestigious award. Award citation: `http://www.sci.utah.edu/news/245-rosenblatt2010.html`

3. Chris Johnson, Visualization Career Award, 2010.

4. Christoph Garth, UC Davis award for Excellence in Postdoctoral Research.

5. Christoph Garth, Hari Krishnan, and Ken Joy. SciDAC 2010 Visualization Night People's Choice Award, "Visualization of Connective Flow with Integral Surfaces."

6. Jamison Daniel, Dave Pugmire, Michael Matheson, and Sean Ahern. SciDAC 2010 Visualization Night People's Choice Award, "Clean Energy for the Future with the ITER Reactor."

7. Hank Childs, Haito Ma, Stan Woosley, John Bell, Ann Almgren, Andy Nonaka. SciDAC 2010 Visualization Night People's Choice Award, "Type Ia Supernova: Turbulent Combustion on the Grandest Scale."

8. Prabhat, Michael Wehner, and E. Wes Bethel. SciDAC 2010 Visualization Night People's Choice Award, "Hurrican Season."